

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 17:54:18 ; Search time 1646.7 Seconds
(without alignments)
5302.013 Million cell updates/sec

Title: US-09-990-099-21

Perfect score: 300
Sequence: 1 attcatcgttcgtatcgtggg.....gattccagagactatgatt 300

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenBml:*

- 1: gb_da:*
- 2: gb_hlg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pt:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vl:*
- 30: em_hlg_hum:*
- 31: em_hlg_inv:*
- 32: em_hlg_other:*
- 33: em_hlg_mus:*
- 34: em_hlg_pln:*
- 35: em_hlg_rod:*
- 36: em_hlg_mam:*
- 37: em_hlg_vrt:*
- 38: em_sy:*
- 39: em_hlg_hum:*
- 40: em_hlg_mus:*
- 41: em_hlg_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	300	100.0	720	1	AF407018
C 2	300	100.0	2886	1	ECOPROTS
C 3	300	100.0	10120	1	AE000446
C 4	300	100.0	136254	2	ECODW82
C 5	300	100.0	252977	2	AC073670
C 6	300	100.0	256377	2	AC020870
C 7	300	100.0	272545	2	AC090533
C 8	298.4	99.5	12642	1	AE005600
C 9	298.4	99.5	280900	1	AP002566
C 10	295	98.3	221285	2	AC023611
C 11	184.4	61.5	245050	1	AL627280
C 12	182.8	60.9	22492	1	AE008877
C 13	36.8	12.3	181977	2	AC103606
C 14	36.6	12.2	160174	2	AP005103
C 15	36.2	12.1	195007	2	CNS06C81
C 16	36	12.0	108822	9	AC004968
C 17	36	12.0	178201	2	AC009169
C 18	36	12.0	207942	9	AC069499
C 19	35	11.7	139534	2	AP005640
C 20	34.6	11.5	1380	10	AF232010
C 21	34.6	11.5	137426	2	AC079357
C 22	34.6	11.5	257214	2	AC126248
C 23	34.6	11.5	299480	2	AC121563
C 24	34.4	11.5	1853	9	AK000917
C 25	34.4	11.5	85413	8	AC009324
C 26	34.4	11.5	153772	9	AC104464
C 27	34.4	11.5	158898	2	AL359879
C 28	34.4	11.5	171482	9	AC034247
C 29	34.4	11.5	171961	2	AC113408
C 30	34.4	11.5	200175	2	AC096004
C 31	34	11.3	78376	9	AL359692
C 32	34	11.3	95836	9	AC016710
C 33	33.8	11.3	128597	2	AC098295
C 34	33.8	11.3	164879	2	AC007876
C 35	33.8	11.3	186747	9	AC018605
C 36	33.8	11.3	199913	9	AC007495
C 37	33.6	11.2	123379	9	AL359715
C 38	33.6	11.2	151040	2	AC068389
C 39	33.6	11.2	161595	2	AC012267
C 40	33.6	11.2	175263	2	AC023102
C 41	33.6	11.2	176856	2	AC124145
C 42	33.4	11.1	98494	9	AC087763
C 43	33.4	11.1	125020	9	AF429315
C 44	33.4	11.1	176884	2	AC083927
C 45	33.4	11.1	185510	2	AC011018

ALIGNMENTS

RESULT 1
AF407018/c
LOCUS AF407018 720 bp DNA linear BCT 10-SEP-2001
DEFINITION Escherichia coli inner membrane protein (ecfI) gene, complete cds.
ACCESSION AF407018
VERSION AF407018.1 GI:15529642
KEYWORDS
SOURCE
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia coli.
REFERENCE
AUTHORS Darligalogue, C., Missiakas, D. and Raina, S.
TITLE Characterization of the Escherichia coli sigma E regulon
JOURNAL J. Biol. Chem. 276 (24), 20866-20875 (2001)

/note="f416; 100 pct identical amino acid sequence and equal length to YDR_ECOLI SW: P31455"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAC76712.1"
 /db_xref="GI:1790124"
 /translation="MAGPVLVYQDRAMKQITFAPRNHLLTNTMTTPSPQWLVDPVRSQSPGTEIERVNIHTGEVEIYRASGAVGVYVTPKSEKVFTHGTPNDETMH YDFHRRGIVIEGKMSNLDMADITAPYRGLRGSHVHFSRGEKVSSTYNDHVA HELDPAIDRVNVAAPCPVNVOKHREISGSHCVLVSRTPTDQSGDEINRAY EGVGYNHAIAPIGITLSPKSKYPELIVLPDDEAKRAGAPLSGTEITLAPR BGVYRRTFTTHRAIPGLVNPVRHWACNPGQIATFLMRDDNGIYOLMLISQGGF VSLTHNTDIQSAPNMPHSGEWLGFVLDNRACAHQSGEVEYTFEHANSPSADAV VSPDQWLAAMEGQWLITETDR"
 2359..2388
 /note="Factor Sigma70; predicted +1 start at 3866948"
 2430..3515
 /gene="yids"
 /note="b3690"
 2430..3515
 /gene="yids"
 /function="orf; Unknown"
 /note="o361; 100 pct identical amino acid sequence and equal length to YIDS_ECOLI SW: P31456"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAC76713.1"
 /db_xref="GI:1790125"
 /translation="MMNVSEKMEHFDVAITIGPAGSALARKAGKQVIALKKHOC GTEGFSKPCGGLAPDAORSEFIDGLDLPDVIANPITFVKTVDAASLITRYQSY INIRHAPDLMMKSLIPASVEYHDSICRTKEDDWHVIFRADGEOHITARYLVG ADGANSVRRHLVDPHOIRKYVAIQOMFAKHPVPEFSCIFDINSITNCYSISKDY FIRGATAPKDGOTREFTLTKKSAEQFOGKVKSECTFSPMODPVCCKDNFV LIGEAQFISASSLEGISYALDSIDILRSVLLAQEKLNTAYVRATRKRLKLFKIV KSRCLTAPALRKIMRSGVAHIFOLKD"
 complement(2436..2464)
 /note="factor Sigma70; predicted +1 start at 3866982"
 complement(2524..2552)
 /note="factor Sigma70; predicted +1 start at 3867070"
 complement(3512..4849)
 /gene="dgot"
 /note="b3691"
 complement(3512..4849)
 /gene="dgot"
 /function="transport; Transport of small molecules: Carbohydrates, organic acids, alcohols"
 /note="f445; formerly designated yidT"
 /codon_start=1
 /transl_table=11
 /product="D-galactonate transport"
 /protein_id="AAC76714.1"
 /db_xref="GI:1790126"
 /translation="MYSGRAMPKIKRKLAMDIPVNAKGRRRYLLVWIFITVICY VBRANLAVASAHIOEFGITKAMGVESAFAMLYLCOIPGMLDRGSRVAFIA ITGWSVALTFEGFATGLMSLIGIRATIFGFAPEPTNNRWMTSEPEHRSAAVFA TSGQVGLAFPLPLIWIQEMLSMHWFIYTGIGIIMSLIMKRYOOPRLTKGSKA ELDYIDGGGLVDGAPVKKAROPLTAKMDKLVHRKILGYLQDRAVASTLWFLT NPPNYLQEKITALKAGMTVPPLAFCVGLSGWADLVLRGFSLGFARKTPII CGLLSTICMGANTNDPMIMCMAALFAGFASITWSLAPMLIGLGGVF NFAGGLGITVPLVYGYLAOGYGFAPALVISAVALIGALSYLLVGVKRRG"
 complement(4924..6687)
 /gene="dgoA"
 /note="b3692"
 complement(4924..6687)
 /gene="dgoA"
 /function="enzyme; Degradation of small molecules: Carbon compounds"
 /note="f587; formerly designated yidU"
 /codon_start=1
 /transl_table=11
 /product="2-oxo-3-deoxygalactonate 6-phosphate aldolase"

and galactonate dehydratase"
 /protein_id="AAC76715.1"
 /db_xref="GI:2367262"
 /translation="MOMQIKLPLILRLITGTPDEALAHGAVIDAFDAVEIPLNSPO MEOSIPALVDYDKALIGAGTVLKPEDVDALARMCCOLIVPINHSEVIRAVGYGM TVPCGATATAEFTALEAGAAAGENISVIGFSAHIVSKSIAIGHRLICMRDRAE NLAOWIDAGCAGAGALSIDLYRAQGSVERFAQQAARFVKAYREAGANKIKITTYRLRP
 Query Match 100.0%; Score 300; DB 1; Length 10120;
 Best Local Similarity 100.0%; Pred No. 5.3e-87;
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATTCATCTGTTGATCGTGGGTTGGCTGATGATTAGCATCCCTGCTGAANA 60
 796 ATTCATCTGTTGATCGTGGGTTGGCTGATGATTAGCATCCCTGCTGAANA 737
 61 ACATCATCATTCATCGTCACGTGGGGGATGCGACTTTAGCTTTCGTCGCCCT 120
 736 ACATCATCATTCATCGTCACGTGGGGGATGCGACTTTAGCTTTCGTCGCCCT 677
 121 CAGCTATGCAATAGACCATTAACCTGCAAAAAAGTCCGCTGATAAGCTTGAAGA 180
 676 CAGCTATGCAATAGACCATTAACCTGCAAAAAAGTCCGCTGATAAGCTTGAAGA 617
 181 CATTCAGACCATTTTACATCGTACCGGATGAGAGCGGCTGATGGCTTGTGCG 240
 616 CATTCAGACCATTTTACATCGTACCGGATGAGAGCGGCTGATGGCTTGTGCG 557
 241 TACCTGACCTGTCATTTGGAAGGCTTACATTCGCTGATTCAGAGCATTAATG 300
 556 TACCTGACCTGTCATTTGGAAGGCTTACATTCGCTGATTCAGAGCATTAATG 497

RESULT 4
 LOCUS ECOLW82/c 136254 bp DNA linear BCT 07-FEB-1995
 DEFINITION E. coli; the region from 81.5 to 84.5 minutes.
 ACCESSION L10328
 VERSION L10328.1 GI:290484
 KEYWORDS
 SOURCE Escherichia coli K12 strain MG1655; lambda clones EC14-52, EC17-187, EC27-890, EC21-99, EC22-175, EC27-236, EC17-30, EC17-137, EC19-71, EC15-15, EC17-159; subclones in M13mp19 or Janus.
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 136254)
 AUTHORS Burland, V., Plunkett, G. III, Daniels, D. L. and Blattner, F. R.
 TITLE DNA sequence and analysis of 136 kilobases of the Escherichia coli genome: organizational symmetry around the origin of replication
 JOURNAL Genomics 16 (3), 551-561 (1993)
 MEDLINE 93315143
 PUBMED 7686882

COMMENT This sequence was determined as part of the E. coli Genome Project (Frederick R. Blattner, director) at the University of Wisconsin-Madison. Supported by award HG00301 from the NIH Human Genome Project. A preliminary report was presented at the Hilton Head meeting Genome Sequencing and Analysis IV, September 26 - 30, 1992. The entire sequence was independently determined from E. coli MG1655; overlaps and conflicts with other sequence determinations are annotated. The end of this entry overlaps the start-of the entry ECOLW85 (M67049) by the six bases of an EcoRI site.

Data kindly submitted in computer readable form by: Guy Plunkett III
 Laboratory of Genetics
 University of Wisconsin
 445 Henry Mall
 Madison WI 53706
 USA
 Phone: 608-262-2534
 Email: ecoligenetics@wisc.edu
 Fax: 608-263-7459

[1] Authors request hold until publication.

```

FEATURES
    source
        location/Qualifiers
            1..136254
            /organism="Escherichia coli"
            /db_xref="taxon:562"
    misc_feature
        1..11607
        /note="EC14-52 (lambda clone)"
        /complement(<1..669)
        /note="X06036; ECORPBG(1..>669)"
        /complement(1..544)
        /gene="fpg"
        /complement(<1..544)
        /gene="fpg"
        /note="alternate gene name mutH"
        /codon_start=1
        /transl_table=1
        /product="Formamidopyrimidine-DNA glycosylase"
        /protein_id="AA61988.1"
        /db_xref="GI:643609"
        /translation="MPPLPEVETSRNGIEPHLVGATILHAHVENGRLMPVSEIYRL
        SDQVLSVORAKYLLLEPEGMWIIHGMGSLRILPEELPEKHDHDLVNSGKV
        LRYTDPFRGAMWTFKEGHNVTLHGPEPLSDDFNGEYLHOKCAKKTAIKPWLMD
        NKLIVGNGNIASSELFAACI"
    misc_feature
        complement(526..1290)
        /note="J01677; ECORPBG(1..764)"
        /complement(557..586)
    promoter
        /note="promoter-like sequence; promoter matrix score of
        44"
    misc_difference
        596..601
        /note="AAAAA in X06036; AAAAAA in J01677 and here"
    misc_difference
        611
        /note="G in J01677; A in X06036 and here"
    misc_difference
        615
        /note="G in J01677; A in X06036 and here"
    misc_difference
        618..623
        /note="TTTTTT in X06036; TTTTTT in J01677 and here"
        /complement(642..809)
        /gene="rpmg (CG Site No. 236)"
        /complement(642..809)
        /gene="rpmg (CG Site No. 236)"
        /codon_start=1
        /transl_table=1
        /product="50S ribosomal subunit protein L33"
        /protein_id="AA61989.1"
        /db_xref="GI:290486"
        /translation="MAKGIREKIKVSSAGTGHPYTTTKNKRKPEKLEKKDFVVR
        QHVIYREAKIK"
    gene
        complement(830..1066)
        /gene="rpmB (CG Site No. 240)"
        /complement(830..1066)
        /gene="rpmB (CG Site No. 240)"
        /codon_start=1
        /transl_table=1
        /product="50S ribosomal subunit protein L28"
        /protein_id="AA61990.1"
        /db_xref="GI:290487"
        /translation="MSRVCOVGTGRPVTCGNRRSHALNATKRRLPLNLSHRVSESEK
        RFTVLRVSAKMVYIDKKGIDYVLAELRARGEKY"
    misc_difference
        902
        /note="G in J01677; A here"
    misc_difference
        1128
        /note="TC in J01677; C here"
    misc_difference
        1153..1155
        /note="T in J01677; TAT here"
    misc_feature
        complement(1168..1819)
        /note="X63366; ECORADC(1..651)"
        /complement(1212..1240)
        /note="promoter-like sequence; promoter matrix score of
        77"
    gene
        complement(1283..1957)
        /gene="radC (CG Site No. 13913)"
        /complement(1283..1957)
        /gene="radC (CG Site No. 13913)"
    CDS

```

```

    misc_difference
        variation
            /note="A deleted in radC102 mutation (X63367)"
            1436
            /note="A deleted in radC102 mutation (X63367)"
            1698..1822
            /note="predicted bend of 76 degrees"
            1875..1902
            /note="promoter-like sequence; promoter matrix score of
            50"
            /complement(1911..1940)
            /gene="radC (CG Site No. 13913)"
            /note="promoter-like sequence; promoter matrix score of
            45"
            2051..3343
            /gene="dtp (CG Site No. 18430)"
            2051..3343
            /gene="dtp (CG Site No. 18430)"
            /codon_start=1
            /transl_table=1
            /protein_id="AA61992.1"
            /db_xref="GI:290489"
            /translation="MKARQKCYCKIANFCHPTGKTIIMSLAGKIKVIGVSGGTAAYK
            PBLVRLRLDRGADVAVAMENAKAFITPLSLQAVSGIPVSDLLDRAAEAKMHTL
            GKWADVLAPATADILARVAGAMANDVSTICLATPAPVAVLPMNOMQRAATQH
            NLEVLASRGLLIMGPSGSOAGDIPGRAXDPLTIVMAVAHFSVPVMDLHLIMT
            AGTREPRLPVYISNHSKMGFAIAAARAGANVILVSGPVSLPTPPKRVNVA
            TALEMBAVYMAVCOONIFIGCAAVADRAATVAPEKIKKOATGODELTITKMNPD
            VAGVAALKDHRPVPVGFAPAEPTNNVEEYRORIRKNDLICANDVSCQTQCFNSDNN
            LHLFWDGDKVLPLEKRELGLDLDELTYTRIDENRR"
            2982..5552
            /note="V01578; ECODUTPYR(1..2568)"
            2982..4591
            /note="X01714; ECOPYREA(1..1609)"
            3324..3379
            /gene="dut (CG Site No. 832)"
            3324..3379
            /gene="dut (CG Site No. 832)"
            /codon_start=1
            /transl_table=1
            /product="deoxyuridine triphosphatase"
            /protein_id="AA61993.1"
            /db_xref="GI:290490"
            /translation="MKRIDVKIILDPYKGFELPYATSGSAGLIDLRACINDAVELAP
            GDTLVPTGLAIHLADPSLAAMMLPSGLGKHGIVLGNVLGLSDSGQGLMISWN
            RGDSFTIOPGERIAQMIFVYVQAEFNLVEDVDATDRBGGRGHSGRO"
            3844..4482
            /gene="tik"
            3844..4482
            /gene="tik"
            /function="unknown"
            /codon_start=1
            /transl_table=1
            /product="23 kD protein"
            /protein_id="AA61994.1"
            /db_xref="GI:290491"
            /translation="MPGKCLFSCVPCNMAEKONAKNRREILQSLALMLSSDGSQ
            RITTKLAASGVSEALYRHPSKTRFSDLSIEFTISITINILIKKEXTARL
            RLIVLILGFERNGLRILITGHALMFEDDRGLRIQLPERIAOLRYLKEKMR
            BEGYTDETILASQILAFCEGMLSRVRSSEFKYRPTDFDARPLTAOLQ"
            4469..4471
            /gene="tik"
            /note="CC in X01714 and V01578; CAC here"
    misc_difference

```

repeat_region 4489..4514
 gene /note="REP (repetitive extragenic palindromic) element;
 contains 1 REP sequence"
 CDS complement(4519..5160)
 /gene="pyrE (CG site No. 327)"
 complement(4519..5160)
 /gene="pyrE (CG site No. 327)"
 /codon_start=1
 /transl_table=1
 /product="Orotate phosphoribosyltransferase"
 /protein_id="AA061995.1"
 /db_xref="GI:643610"

/translation="MKPYQKQFLEPALKSVKLVKFEPTLKGKRSYFENAGLENTGR
 DLALGRTFAIDLVDSIEFDLFGPAKGIPIATTTAVALAHHDLDPCNKEA
 KDHEGSLVGSALQGVMLVDVITAGTAIRSEMETIOANGHDLIAGVILISIDROERG
 RGETSAIOEYERDYNCKVISIITLKDLIAVLEKPEMAELAAVKVIREFVY"

Query Match 100.0%; Score 300; DB 1; Length 136254;
 Best Local Similarity 100.0%; Pred No. 6.8e-87;
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATTATCTGTGATCGTGGGTGTTGGCCTGATGATGATTAAGCATCCCTGCGAAATA 60
 Db 57116 ATTATCTGTGATCGTGGGTGTTGGCCTGATGATGATTAAGCATCCCTGCGAAATA 57057
 OY 61 ACATCATCATTAAGTCCGACTGTGGCGGTATGCGACTTAAGCTTGGTGGTCCCT 120
 Db 57056 ACATCATCATTAAGTCCGACTGTGGCGGTATGCGACTTAAGCTTGGTGGTCCCT 56997
 OY 121 CAGCTATGCAATGACCAATAAAGTCCGCTGATTAAGCTTGAAGTT 180
 Db 56996 CAGCTATGCAATGACCAATAAAGTCCGCTGATTAAGCTTGAAGTT 56937
 OY 181 CATTTCCAGACCATTTTTCATGTCGATGAGGAGCGCGCTGATGCTGTTGTCG 240
 Db 56936 CATTTCCAGACCATTTTTCATGTCGATGAGGAGCGCGCTGATGCTGTTGTCG 56877
 OY 241 TACCTGACCTGTCATGTCGATGAGGAGCTTACATTTCCGCTGATTAAGCTTGAAGTT 300
 Db 56876 TACCTGACCTGTCATGTCGATGAGGAGCTTACATTTCCGCTGATTAAGCTTGAAGTT 56817

RESULT 5
 AC073670/c 252977 bp DNA linear HTG 08-AUG-2000
 LOCUS AC073670
 DEFINITION Mus musculus clone CT7-368A6, WORKING DRAFT SEQUENCE, 85 unordered
 pieces.
 AC073670
 AC073670.1 GI:8810287
 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS Mus musculus.
 SOURCE Mus musculus.
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 252977)
 AUTHORS DOE Joint Genome Institute.
 TITLE Sequencing of Mouse
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 252977)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 COMMENT -----Genome Center
 Center: Joint Genome Institute
 Center Code: JGI
 Web site: http://www.jgi.doe.gov

Project Information
 Center Project Name: 1094499
 Center clone name: RG-MBAC_368A6
 Summary Statistics

Consensus quality: 177673 bases at least Q40
 Consensus quality: 214946 bases at least Q30
 Consensus quality: 224156 bases at least Q20
 Estimated insert size: 100000; pulse field gel estimation
 Estimated insert size: 244577; sum-of-contigs estimation
 Quality coverage: 6.6 in Q20 bases; pulse field gel estimation
 Quality coverage: 2.7 in Q20 bases; sum-of-contigs estimation
 NOTE: This is a 'working draft' sequence. It currently
 consists of 85 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1	1111	1110: contig of 1110 bp in length
*	1211	1210: gap of unknown length
*	1211	1210: gap of unknown length
*	2458	2457: contig of 1247 bp in length
*	2538	2537: gap of unknown length
*	3664	3663: contig of 1106 bp in length
*	3764	3763: gap of unknown length
*	4897	4896: contig of 1133 bp in length
*	4997	4996: gap of unknown length
*	6153	6152: contig of 1156 bp in length
*	6253	6252: gap of unknown length
*	7288	7288: contig of 1036 bp in length
*	7389	7388: gap of unknown length
*	8965	8964: contig of 1576 bp in length
*	9065	9064: gap of unknown length
*	10175	10174: contig of 1110 bp in length
*	10275	10274: gap of unknown length
*	11587	11586: contig of 1312 bp in length
*	11687	11686: gap of unknown length
*	12874	12873: contig of 1187 bp in length
*	12974	12973: gap of unknown length
*	14205	14204: contig of 1231 bp in length
*	14305	14304: gap of unknown length
*	15842	15841: contig of 1537 bp in length
*	15942	15941: gap of unknown length
*	17159	17159: contig of 1218 bp in length
*	17259	17259: gap of unknown length
*	18555	18555: gap of unknown length
*	18656	18655: gap of 1296 bp in length
*	19742	19742: gap of unknown length
*	19842	19842: contig of 1087 bp in length
*	19843	19842: gap of unknown length
*	21274	21273: contig of 1431 bp in length
*	21374	21373: gap of unknown length
*	22751	22750: contig of 1377 bp in length
*	22851	22850: gap of unknown length
*	24001	24000: gap of 1150 bp in length
*	24101	24100: gap of unknown length
*	25538	25538: contig of 1438 bp in length
*	25638	25638: gap of unknown length
*	26726	26726: contig of 1088 bp in length
*	26827	26826: gap of unknown length
*	28012	28011: contig of 1185 bp in length
*	28112	28111: gap of unknown length
*	29531	29531: contig of 1420 bp in length
*	29632	29631: gap of unknown length
*	31152	31152: contig of 1521 bp in length
*	31253	31252: gap of unknown length
*	32452	32452: contig of 1200 bp in length
*	32453	32452: gap of unknown length
*	32553	32552: gap of unknown length
*	33655	33654: contig of 1102 bp in length
*	33755	33754: gap of unknown length
*	34775	34774: contig of 1020 bp in length
*	34875	34874: gap of unknown length
*	36052	36051: contig of 1177 bp in length
*	36152	36151: gap of unknown length
*	37591	37590: contig of 1439 bp in length
*	37691	37690: gap of unknown length
*	38908	38908: contig of 1218 bp in length
*	38909	38908: gap of unknown length

*	39009	40601:	gap of 1593 bp in length
*	40602	40701:	gap of unknown length
*	40702	41926:	contig of 1225 bp in length
*	41927	42026:	gap of unknown length
*	42027	43993:	contig of 1957 bp in length
*	43984	44083:	gap of unknown length
*	44084	45206:	contig of 1123 bp in length
*	45207	45306:	gap of unknown length
*	45307	46881:	contig of 1575 bp in length
*	46882	46981:	gap of unknown length
*	46982	48056:	contig of 1075 bp in length
*	48057	48156:	gap of unknown length
*	48157	49666:	contig of 1512 bp in length
*	49669	49768:	gap of unknown length
*	49769	51187:	contig of 1419 bp in length
*	51188	51287:	gap of unknown length
*	51288	52332:	contig of 1045 bp in length
*	52333	52432:	gap of unknown length
*	52433	53755:	contig of 1322 bp in length
*	53755	53854:	gap of unknown length
*	53855	55581:	contig of 1127 bp in length
*	55582	55681:	gap of unknown length
*	55682	57262:	contig of 1581 bp in length
*	57263	57362:	gap of unknown length
*	57363	59152:	contig of 1790 bp in length
*	59153	59252:	gap of unknown length
*	59253	60732:	contig of 1480 bp in length
*	60733	60832:	gap of unknown length
*	60833	63030:	contig of 2198 bp in length
*	63031	63130:	gap of unknown length
*	63131	65366:	contig of 2216 bp in length
*	65367	65466:	gap of unknown length
*	65467	67227:	contig of 1761 bp in length
*	67228	67327:	gap of unknown length
*	67328	68506:	contig of 1119 bp in length
*	68507	68606:	gap of unknown length
*	68607	71259:	contig of 2653 bp in length
*	71260	71359:	gap of unknown length
*	71360	73965:	contig of 2608 bp in length
*	73966	74065:	gap of unknown length
*	74066	76125:	contig of 2060 bp in length
*	76126	76225:	gap of unknown length
*	76226	77297:	contig of 1072 bp in length
*	77298	77397:	gap of unknown length
*	77398	79646:	contig of 2249 bp in length
*	79647	79746:	gap of unknown length
*	79747	82449:	contig of 2703 bp in length
*	82450	82549:	gap of unknown length
*	82550	85345:	contig of 2796 bp in length
*	85346	85445:	gap of unknown length
*	85446	86817:	contig of 1372 bp in length
*	86818	86917:	gap of unknown length
*	86919	90145:	contig of 3228 bp in length
*	90146	90245:	gap of unknown length
*	90246	91490:	contig of 1245 bp in length
*	91491	91590:	gap of unknown length
*	91591	94935:	contig of 3345 bp in length
*	94936	95035:	gap of unknown length
*	95036	97845:	contig of 2810 bp in length
*	97846	97945:	gap of unknown length
*	97946	100449:	contig of 2504 bp in length
*	100450	100549:	gap of unknown length
*	100550	103583:	contig of 3034 bp in length
*	103584	103683:	gap of unknown length
*	103684	107672:	contig of 3995 bp in length
*	107673	107778:	gap of unknown length
*	107779	110534:	contig of 2756 bp in length
*	110535	110634:	gap of unknown length
*	110635	113491:	contig of 2857 bp in length
*	113492	113591:	gap of unknown length
*	113592	115603:	contig of 2012 bp in length
*	115604	118957:	gap of unknown length
*	118958	119570:	contig of 3254 bp in length

Query Match	Best Local	Similarity	100.0%;	Score 300;	DB 2;	Length 252977;
Matches	300;	Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;
QY 1	ATTCATCGTGTGATCGGGGCTGTGGCCTGATGATATAGCATCCCTTCTGTAATA	60				
Db 11269	ATTATCTGTTGATCGTGGGCTGTGGCCTGATGATATAGCATCCCTTCTGTAATA	11210				
QY 61	ACATCATCATTAAGCTGCGACATGTGGCGGCTATGCGACATTTAAAGCTTGTCTGCCCCCT	120				
Db 11209	ACATCATCATTAAGCTGCGACATGTGGCGGCTATGCGACATTTAAAGCTTGTCTGCCCCCT	11150				
QY 121	CAGCTATGCAATGAGCATTAACATGCAAAAAAGTCCGCTGATTAAGCTTGAAGATT	180				
Db 11149	CAGCTATGCAATGAGCATTAACATGCAAAAAAGTCCGCTGATTAAGCTTGAAGATT	11090				
QY 181	CATTTCCAGACCCATTTTACATCGCTAGCCGATGAGACGCGCCTGATGCGTCTCTGCGC	240				
Db 11089	CATTTCCAGACCCATTTTACATCGCTAGCCGATGAGACGCGCCTGATGCGTCTCTGCGC	11030				
QY 241	TACCTGACCTGTCCTCATTTGTGGAAGGCTTACATTCCTGCTGATTCAGAGCTATTGATT	300				
Db 11029	TACCTGACCTGTCCTCATTTGTGGAAGGCTTACATTCCTGCTGATTCAGAGCTATTGATT	10970				
RESULT 6						
AC020870/c	AC020870	256373 bp	DNA	linear	HTG 17-FEB-2000	
LOCUS	DEFINITION	Mus musculus clone RP23-302J15, LOW-PASS SEQUENCE SAMPLING.				
AC020870	AC020870					
VERSION	AC020870.2	GI:6984372				
KEYWORDS	HTG; HTGS_PHASE0.					
SOURCE	Mus musculus.					
ORGANISM	Mus musculus					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
TITLE	1 (bases 1 to 256373)					
JOURNAL	DOE Joint Genome Institute.					
AUTHORS	Unpublished					
TITLE	2 (bases 1 to 256373)					
JOURNAL	DOE Joint Genome Institute.					
AUTHORS	Direct Submission					
TITLE	Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint					
JOURNAL	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA					
AUTHORS	On Feb 17, 2000 this sequence version replaced gi:6686438.					
TITLE	* NOTE: This record contains 198 individual					

* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 752: contig of 752 bp in length
* gap of unknown length
* 753 1000: contig of 248 bp in length
* gap of unknown length
* 1001 1604: contig of 604 bp in length
* gap of unknown length
* 1605 2302: contig of 698 bp in length
* gap of unknown length
* 2303 2699: contig of 397 bp in length
* gap of unknown length
* 2700 3686: contig of 987 bp in length
* gap of unknown length
* 3687 4099: contig of 413 bp in length
* gap of unknown length
* 4100 4778: contig of 679 bp in length
* gap of unknown length
* 4779 5571: contig of 793 bp in length
* gap of unknown length
* 5572 5967: contig of 396 bp in length
* gap of unknown length
* 5968 6763: contig of 796 bp in length
* gap of unknown length
* 6764 7639: contig of 876 bp in length
* gap of unknown length
* 7640 8442: contig of 803 bp in length
* gap of unknown length
* 8443 9414: contig of 972 bp in length
* gap of unknown length
* 9415 10236: contig of 822 bp in length
* gap of unknown length
* 10237 10844: contig of 608 bp in length
* gap of unknown length
* 10845 11504: contig of 660 bp in length
* gap of unknown length
* 11505 12321: contig of 817 bp in length
* gap of unknown length
* 12322 12652: contig of 331 bp in length
* gap of unknown length
* 12653 14116: contig of 1464 bp in length
* gap of unknown length
* 14117 15298: contig of 1182 bp in length
* gap of unknown length
* 15299 16234: contig of 936 bp in length
* gap of unknown length
* 16235 16376: contig of 142 bp in length
* gap of unknown length
* 16377 16664: contig of 288 bp in length
* gap of unknown length
* 16665 17631: contig of 967 bp in length
* gap of unknown length
* 17632 19241: contig of 1610 bp in length
* gap of unknown length
* 19242 20453: contig of 1212 bp in length
* gap of unknown length
* 20454 20559: contig of 106 bp in length
* gap of unknown length
* 20560 21117: contig of 558 bp in length
* gap of unknown length
* 21118 21655: contig of 538 bp in length
* gap of unknown length
* 21656 21818: contig of 163 bp in length
* gap of unknown length
* 21819 22679: contig of 861 bp in length
* gap of unknown length

* 22680 23579: gap of unknown length
* contig of 900 bp in length
* 23580 24383: gap of unknown length
* contig of 804 bp in length
* 24384 25612: gap of unknown length
* contig of 1229 bp in length
* 25613 25957: gap of unknown length
* contig of 345 bp in length
* 25958 27260: gap of unknown length
* contig of 1303 bp in length
* 27261 28596: gap of unknown length
* contig of 1336 bp in length
* 28597 29473: gap of unknown length
* contig of 877 bp in length
* 29474 30372: gap of unknown length
* contig of 899 bp in length
* 30373 31202: gap of unknown length
* contig of 830 bp in length
* 31203 32688: gap of unknown length
* contig of 1486 bp in length
* 32689 33431: gap of unknown length
* contig of 743 bp in length
* 33432 33693: gap of unknown length
* contig of 262 bp in length
* 33694 34616: gap of unknown length
* contig of 923 bp in length
* 34617 35884: gap of unknown length
* contig of 1268 bp in length
* 35885 36459: gap of unknown length
* contig of 575 bp in length
* 36460 36996: gap of unknown length
* contig of 537 bp in length
* 36997 37941: gap of unknown length
* contig of 945 bp in length
* 37942 38682: gap of unknown length
* contig of 741 bp in length
* 38683 39549: gap of unknown length
* contig of 867 bp in length
* 39550 40792: gap of unknown length
* contig of 1243 bp in length
* 40793 42585: gap of unknown length
* contig of 1793 bp in length
* 42586 44115: gap of unknown length
* contig of 1530 bp in length
* 44116 45300: gap of unknown length
* contig of 1185 bp in length
* 45301 46417: gap of unknown length
* contig of 1117 bp in length
* 46418 47815: gap of unknown length
* contig of 1398 bp in length
* 47816 49310: gap of unknown length
* contig of 1495 bp in length
* 49311 50304: gap of unknown length
* contig of 994 bp in length
* 50305 50848: gap of unknown length
* contig of 544 bp in length
* 50849 51828: gap of unknown length
* contig of 980 bp in length
* 51829 52699: gap of unknown length
* contig of 871 bp in length
* 52700 53365: gap of unknown length
* contig of 666 bp in length
* 53366 54633: gap of unknown length
* contig of 1268 bp in length
* 54634 55470: gap of unknown length
* contig of 837 bp in length
* 55471 56358: gap of unknown length
* contig of 888 bp in length
* 56359 56934: gap of unknown length
* contig of 576 bp in length
* 56935 57672: gap of unknown length
* contig of 738 bp in length


```

* 230383 236251: contig of 5869 bp in length
* 236252 236271: gap of unknown length
* 240572 240572: contig of 4301 bp in length
* 240573 240582: gap of unknown length
* 240593 242789: contig of 2197 bp in length
* 242790 242809: gap of unknown length
* 242810 244696 244695: contig of 1886 bp in length
* 244696 244715: gap of unknown length
* 244716 246969: contig of 2254 bp in length
* 246970 247890: gap of unknown length
* 247891 247890: contig of 881 bp in length
* 248194 248193: contig of 303 bp in length
* 248214 250122: gap of unknown length
* 250123 250142: contig of 1909 bp in length
* 250143 251367: gap of unknown length
* 251368 251387: contig of 1225 bp in length
* 251388 252749: gap of unknown length
* 252750 252769: contig of 1362 bp in length
* 252770 253660: gap of unknown length
* 253661 253660: contig of 871 bp in length
* 253662 255032: gap of unknown length
* 255033 255052: contig of 1372 bp in length
* 257334 257333: gap of unknown length
* 257354 259409: contig of 2281 bp in length
* 259410 259429: gap of unknown length
* 259430 260459: contig of 2056 bp in length
* 260460 260479: gap of unknown length
* 260480 260829: contig of 1030 bp in length
* 260830 260829: gap of unknown length
* 260831 261661: gap of 350 bp in length
* 261662 261661: contig of 812 bp in length
* 261682 262532: gap of unknown length
* 262533 262552: contig of 851 bp in length
* 263353 263351: gap of unknown length
* 263352 263371: contig of 799 bp in length
* 263372 264217: gap of unknown length
* 264218 264237: gap of 846 bp in length
* 264238 264205: contig of unknown length
* 265206 265225: gap of 968 bp in length
* 265226 266270: gap of unknown length
* 266271 266290: contig of 1045 bp in length
* 266291 266749: gap of unknown length
* 266750 266769: gap of 459 bp in length
* 267719 267718: contig of 949 bp in length
* 267739 268576: gap of unknown length
* 268577 268596: contig of 838 bp in length
* 268597 270010: gap of unknown length
* 270011 270030: contig of 1414 bp in length
* 270031 271654: gap of unknown length
* 271655 271674: contig of 1624 bp in length
* 271675 272545: gap of unknown length
* 272545 272545: contig of 871 bp in length.
FEATURES
source
1. 272545
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RP23-33H13"
/sex="male"
1. 30267
/misc_feature
/note="assembly_name:Contig108"
30288. 58400
/note="assembly_name:Contig107"
clone_end:5p6
vector_side:left"
58421. 88864
/misc_feature
/note="assembly_name:Contig106"
88885. 109889
/misc_feature
/note="assembly_name:Contig105"
109910. 125133

```

```

/misc_feature
/note="assembly_name:Contig104"
clone_end:77
vector_side:right"
125154. 139549
/misc_feature
/note="assembly_name:Contig103"
139570. 153410
/misc_feature
/note="assembly_name:Contig102"
153431. 162567
/misc_feature
/note="assembly_name:Contig101"
162588. 171044
/misc_feature
/note="assembly_name:Contig100"
171065. 179304
/misc_feature
/note="assembly_name:Contig99"
179325. 187059
/misc_feature
/note="assembly_name:Contig98"
187080. 197598
/misc_feature
/note="assembly_name:Contig97"
197619. 205966
/misc_feature
/note="assembly_name:Contig96"
205987. 212449
/misc_feature
/note="assembly_name:Contig95"
212470. 217752
/misc_feature
/note="assembly_name:Contig94"
217773. 223943
/misc_feature
/note="assembly_name:Contig93"
223964. 230362
/misc_feature
/note="assembly_name:Contig92"
230383. 236251
/misc_feature
/note="assembly_name:Contig91"
236272. 240572
/misc_feature
/note="assembly_name:Contig90"
240593. 242789
/misc_feature
/note="assembly_name:Contig89"
242810. 244695
/misc_feature
/note="assembly_name:Contig88"
244716. 246969
/misc_feature
/note="assembly_name:Contig87"
246990. 247890
/misc_feature
/note="assembly_name:Contig86"
247891. 248193
/misc_feature
/note="assembly_name:Contig85"
248214. 250122
/misc_feature
/note="assembly_name:Contig84"

```

```

Query Match
Best Local Similarity 100.0%; Score 300; DB 2; Length 272545;
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ATTCATCTGTGATCTGCTGGCTTGGCCGATGAGTATAGCATCCCTGCTGAAATA 60
Db 257005 ATTCACTGTGTGATCGTGGCTTGGCCGATGAGTATAGCATCCCTGCTGAAATA 257064
QY 61 ACATCATCTGATCGTGGCTTGGCCGATGAGTATAGCATCCCTGCTGAAATA 257064
Db 257065 ACATCATCTGATCGTGGCTTGGCCGATGAGTATAGCATCCCTGCTGAAATA 257124
QY 121 CAGTCTAGCAATAGACCAATTAAGCAAAAAAGCCGCTGTAAGGCTTAAAGTT 180
Db 257125 CAGTCTATGCAATAGACCAATTAAGCAAAAAAGCCGCTGTAAGGCTTAAAGTT 257244
QY 181 CATTTCCAGACCAATTTTACATGTCGCGGATGAGAGCGCGCTGATGCTTGGC 257184
Db 257185 CATTTCCAGACCAATTTTACATGTCGCGGATGAGAGCGCGCTGATGCTTGGC 257244
QY 241 TACCTGACGTCGATCTTGTGAAGGTCTTACATTCGCGTGAATTCAGAGCTATGATT 300
Db 257245 TACCTGACGTCGATCTTGTGAAGGTCTTACATTCGCGTGAATTCAGAGCTATGATT 257304

```

```

RESULT 8
AE005600/c 12642 bp DNA linear BCT 21-MAR-2001
LOCUS Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 219
DEFINITION

```

of 290.
 AEO05600.1 GI:12518521
 Escherichia coli O157:H7 EDL933.
 Escherichia coli O157:H7 EDL933
 Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE
 AUTHORS
 1 (bases 1 to 12642)
 Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D.,
 Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,
 Postel, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,
 Grobleck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K.,
 Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,
 Welch, R.A. and Blattner, F.R.
 Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
 Nature 409 (6819), 529-533 (2001)

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 21074935
 11206551

REFERENCE
 AUTHORS
 2 (bases 1 to 12642)
 Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D.,
 Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,
 Postel, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,
 Grobleck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K.,
 Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,
 Welch, R.A. and Blattner, F.R.
 Direct Submission
 Submitted (22-OCT-2000) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 Location/Qualifiers

FEATURES
 SOURCE
 1.12642
 /organism="Escherichia coli O157:H7 EDL933"
 /strain="EDL933"
 /serotype="O157:H7"
 /db_xref="taxon:155864"
 /note="enterohemorrhagic"
 99.1022
 /gene="yldP"
 /note="25175"
 99.1022
 /gene="yldP"
 /note="25175"
 99.1022
 /gene="yldP"
 /function="putative regulator; Not classified"
 /note="Residues 1 to 307 of 307 are 99.34 pct identical to
 residues 1 to 307 of 307 from Escherichia coli K-12 strain
 MG1655: B3680"
 99.1022
 /codon_start=1
 /transl_table=11
 /product="putative ARAC-type regulatory protein"
 /protein_id="AAG58883.1"
 /db_xref="GI:12518522"
 /translation="MSIIRRDIKKMGKLGTSVDYKENTPYNIPILLINENYSSGISI
 SLMTYADENIRYIMPRDKKPLANSWAVYVGGCGKILKNEQITLHNCITFL
 KPMIDHSYHCEGLWEDQYMETPTSPMDPIGOOSVYNGEIVNOELTEVALITSP
 EAIKNNLAVFLFKIILYOWICMLKADGKDPORQIEKILATILHRSLOQMSVADMA
 TIPCSEAMLRRLRYTGKTPKRYIDARLDLALSLKQGSVGEVADTLNFDSPH
 FSKAFKHFGYASAVLKNTDHPDASPHN"
 complement(1019..2341)
 /gene="25177"
 complement(1019..2341)
 /gene="25177"
 /function="putative enzyme; Transport of small molecules;
 Carbohydrates, organic acids, alcohols"
 /note="Residues 1 to 440 of 440 are 75.05 pct identical to
 residues 1 to 441 of 441 from Genpept 118 :
 g1|21451521|gb|AAB63015.1| (U81185) Malt (Fusobacterium
 mortiferum)"
 /codon_start=1
 /transl_table=11
 /product="putative 6-phospho-beta-glucosidase"
 /protein_id="AAG58884.1"
 /db_xref="GI:12518523"
 /translation="MTKFSVVAVGGGTFPTQIVYMLLANNODRPLRLKPYDNGSR
 DEVIACACVILKEKAPDIAFSYTTDPVAVSDVDFVMAHTRVCKYIMRELDKPIPLR

misc-feature
 1241..1748
 /note="O-island #151: Region of the EDL933 chromosome not
 homologous to E. coli K-12 MG1655"
 complement(2341..3963)
 /gene="25178"
 complement(2341..3963)
 /gene="25178"
 /function="putative enzyme; Transport of small molecules;
 Carbohydrates, organic acids, alcohols"
 /note="Residues 1 to 370 of 540 are 99.45 pct identical to
 residues 86 to 455 of 455 from Escherichia coli K-12
 strain MG1655: B3683"
 /codon_start=1
 /transl_table=11
 /product="putative PTS component"
 /protein_id="AAG58885.1"
 /db_xref="GI:12518524"
 /translation="MEMLSQIQRFQGMFTPVLPFPACTYVGLAIIIONPFEVGS
 TDPSLRQAVIIEEGGTVFPRNPLIFAVGLPIGLAKQAGACLAVMGSLTWN
 FIMMGWMSRGVDPDQAVAGSGITLMAAGIKLTDTSIIAIIISGIVLRHRL
 DKLPVGLIETGSIYVITAFIWMIPCAVLTLGMPKQOMQIESLSAEPKLEPE
 WYTFLERLILPTGLHFTITGPFPAGAVEGGIOWMHOIESLSAEPKLEPE
 GGFALHNSKIFGAVGISLAFPAIENRNVKAVGLIPATLAVGTLEPTEPL
 FTSPLFAVHVLAAASVWYLYGVNNGGLIDQVLPONNIPMSHADMLOI
 AIGCFLLYEVFRTLLDPMCTPGREDAEVLKSKAEYKASRCOTFAAPKELD
 QAAQTLQAGGVGSSINNCATRIALRILADHDSQDDEVFKLGHGFRSGDAIOV
 ILGHVSQLRQDLSINSQSAENVAITEAV"
 4253..4969
 /gene="yldP"
 /note="25179"
 4253..4969
 /gene="yldP"
 /function="putative regulator; Not classified"
 /note="Residues 1 to 238 of 238 are 99.57 pct identical to
 residues 1 to 238 of 238 from Escherichia coli K-12 strain
 MG1655: B3684"
 /codon_start=1
 /transl_table=11
 /product="putative transcriptional regulator"
 /protein_id="AAG58886.1"
 /db_xref="GI:12518525"
 /translation="MYKSTAEKRLRLNSADFTLNSLPGEKKLAEEFVSRMTIRK
 AIDLAVAGLAVRRHSGTYLVKRDVHOTFASITGLVEVLRKQGTVTSOVLIEIMP
 APPAIALSRQIINEQIYFSRVRFPVGRKPLMEDSYMVKLFRLMSLOHESKFEY
 IEQCGILIGNEISLPVADRLAROMKVAETPLTLITSLYSESEFPLNSVMF
 RNASEYQVEYHLRLHPDKS"
 complement(4966..6651)
 /gene="yldP"
 complement(4966..6651)
 /gene="yldP"
 /function="putative transport; Not classified"
 /note="Residues 1 to 561 of 561 are 100.00 pct identical
 to residues 1 to 561 of 561 from Escherichia coli K-12
 strain MG1655: B3685"
 /codon_start=1
 /transl_table=11
 /product="putative transport protein"
 /protein_id="AAG58887.1"
 /db_xref="GI:12518526"
 /translation="MLOSQKMTMSDIALTVSILAVAVGLFIGNVFRFGIGLIGV
 LFGITIGHVVSQAGMTLSMDLHVIOEGLIFVYTGIOVGEFFRSLSAVSRLRN
 LFVGLIVITGDLVTAIHLKLPDILPVGAVTIPGAVTPALGACQQLRLDUGPHEM
 VDQMGMSYAAVFPGLIGLITMMMLRVIFERNVNEAOHESSTNGALIKITINR
 VENENLHDLAKDPLINDGKLTCSLKRENTLKVPSPPTIIOGLDGLAVGQPDH
 NAOVLIGQEVDTSLSTKGDULRVERVYVNEVNLKIRIDLFKFRYVIVSRLNAG
 VELVAGSDISLQFDIILNIGRPSAIDAVANVIGNOQLQGVMLPVFIGIGLVLL

GSIPFVPEGPAAIKLIGLAGSPLIALILIGRISIGKLYMPESPANIALRELIGIVF
LSVYGLKSGGDFVNTLVNGEXLSWIGALITVAVLITVIGILAMLKMNITMCGHL
ASMDPPLPALANLHPISGAALSYATVPLVWFLIIRIPOLLAVLENSIG
/gene="lbp3"
/note="25182"
/complement(6823..7257)

CDS

/function="factor: Adaptations, atypical conditions"
/note="Residues 1 to 144 of 144 are 100.00 pct identical
strain MG1655: B3686"
/codon_start=1
/transl_table=1
/product="heat shock protein"
/protein_id="AAG58888.1"
/db_xref="GI:12518527"
/translation="MTMRNFDLSPLMRONIGFDKLANALONAGESOSFPYNIKESD
ENMEVSGATFVNLLHITLIRNEPIAORIAISERPALNS"
/gene="lbp3"
/note="25183"
/complement(7363..7776)

CDS

/function="factor: Adaptations, atypical conditions"
/note="Residues 1 to 137 of 137 are 100.00 pct identical
strain MG1655: B3687"
/codon_start=1
/transl_table=1
/product="heat shock protein"
/protein_id="AAG58889.1"
/db_xref="GI:12518528"
/translation="MRNFDLSPLMRONIGFDKLANALONAGESOSFPYNIKESD
ENHIVKAVAGFAESELITADNLVVKAGADEKERTYIYOGIAERNFKEKOLA
8007..8414
/gene="yidQ"
/note="25184"
/complement(8007..8414)

gene

/function="orf: Unknown function"
/note="Residues 1 to 135 of 135 are 99.25 pct identical to
MG1655: B3688"
/codon_start=1
/transl_table=1
/protein_id="AAG58890.1"
/db_xref="GI:12518529"
/translation="MMMLFASADRYNSSGQHPRSTDEFIMIRNVLAIPMISGMLIG
GSSVMSHRTGREGYPTGRASATMIGDFTNMCKSLALIDMFTVMDTLPLMDV

Query Match
Best local Similarity 99.7%: Score 298.4; DB 1: Length 12642;
Matches 299; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATTCACTGTTGATCGTGGGTGGCTGATGAGTTATAGCGATCCCTGCTGAATA 60
DB 8076 ATTCACTGTTGATCGTGGGTGGCTGATGAGTTATAGCGATCCCTGCTGAATA 60
QY 61 ACATCATATTAACGTCGACGTCGCGGCTATCCACTTAAGTTCCGCGGCCCT 120
DB 8016 ACATCATATTAACGTCGACGTCGCGGCTATCCACTTAAGTTCCGCGGCCCT 120
QY 121 CAGTCATGCAATAGACCAATAACTGCAAAAAAAGTCGCGCTATAAGCTTGAAGTT 180
DB 7956 CAGTCATGCAATAGACCAATAACTGCAAAAAAAGTCGCGCTATAAGCTTGAAGTT 180
QY 181 CATTTCAGAGCCATTTTTCATCGTAGCCGATGAGAGCGCGCTGAGGCTTCTGCG 240
DB 7996 CATTTCAGAGCCATTTTTCATCGTAGCCGATGAGAGCGCGCTGAGGCTTCTGCG 240

QY 241 TACCTGACCGTCGATTTGGAGGCTTTACATTCCTGCTGATTTCCAGAGCTATGATT 300
DB 7836 TACCTGACCGTCGATTTGGAGGCTTTACATTCCTGCTGATTTCCAGAGCTATGATT 300

RESULT 9
AP002566/c 280900 bp DNA linear BCT 07-MAR-2001
LOCUS Escherichia coli O157:H7 DNA, complete genome, section 17/20.
DEFINITION AP002566 BA000007
ACCESSION AP002566.1 GI:13363930
VERSION
KEYWORDS
SOURCE
Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
DNA.
Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

ORGANISM
REFERENCE
AUTHORS
Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C.H., Kimura, S.,
Yamamoto, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T.,
Sasakawa, C., and Shinagawa, H.
Complete nucleotide sequence of the prophage VT2-Sakai carrying the
derived from the Sakai outbreak
genes Genet. Syst. 74 (5), 227-239 (1999)
20198760
2 (sites)
REFERENCE
AUTHORS
Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M.,
Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H.
Comparative analysis of the whole set of RNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
20557356
3 (sites)
REFERENCE
AUTHORS
Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S.,
Yutsudo, C.H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T.,
Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C., and
Shinagawa, H.
Complete nucleotide sequence of the prophage VT1-Sakai carrying the
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
20564182
4 (sites)
REFERENCE
AUTHORS
Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K.,
Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T.,
Ogasawara, N., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C.,
Shinagawa, H.
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
21156231
5 (bases 1 to 280900)
REFERENCE
AUTHORS
Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and
Hayashi, T.
Direct Submission
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center: 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail: ken@gen-info.osaka-u.ac.jp,
url: http://www.gen-info.osaka-u.ac.jp/
Fax: 81-6-6879-2047)
genome project.

COMMENT
FEATURES
SOURCE

Location/Qualifiers
1..280900
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
complement(143..616)

gene

QY 121 CAGCTATGCAATAGACATTAAGTCAAAAAAGCTCCGCTGATTAAGCTTGAAGT 180
Db 171537 CAGCTATGCAATAGACATTAAGTCAAAAAAGCTCCGCTGATTAAGCTTGAAGT 171478
QY 181 CATTTCGACACCATTTTACATCGTACCGATGAGCGCGCTGATGGTGTCTGGC 240
Db 171477 CATTTCGACACCATTTTACATCGTACCGATGAGCGCGCTGATGGTGTCTGGC 240
QY 241 TACCTGACCTGTCATTTGGAAGTCTTACATTCGCTGATTTGAGAGCATTTGATT 300
Db 171417 TACCTGACCTGTCATTTGGAAGTCTTACATTCGCTGATTTGAGAGCATTTGATT 171358
RESULT 10
LOCUS AC023611
DEFINITION Mus musculus clone CT7-9K21, WORKING DRAFT SEQUENCE, 62 unordered
pieces.
ACCESSION AC023611 221285 bp DNA linear HTG 29-JUN-2000
VERSION AC023611.2 GI:8810277
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 221285)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 221285)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT -----Genome Center
On Jun 29, 2000 this sequence version replaced gi:6980221.
Center: Joint Genome Institute
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 956898
Center Clone Name: RG-MBAC_9K21

Summary Statistics
Consensus quality: 172196 bases at least Q40
Consensus quality: 196011 bases at least Q20
Estimated insert size: 201702 bases at least Q20
Estimated insert size: 215185; sum-of-contigs estimation
Quality coverage: 6.26 in Q20 bases; pulse field gel estimation
Quality coverage: 4.65 in Q20 bases; sum-of-contigs estimation
NOTE: This is a 'working draft' sequence. It currently
* consists of 62 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1033: contig of 1033 bp in length
* 1034 1033: gap of unknown length
* 1134 1133: gap of unknown length
* 2542 2542: contig of 1409 bp in length
* 2643 2642: gap of unknown length
* 3698 3698: contig of 1056 bp in length
* 3799 3798: gap of unknown length
* 4993 4993: contig of 1195 bp in length
* 4994 5093: gap of unknown length
* 5094 6427: contig of 1334 bp in length
* 6428 6527: gap of unknown length
* 7763 7763: contig of 1236 bp in length
* 7764 7863: gap of unknown length
*
* 7864 8877: contig of 1014 bp in length
* 8878 8877: gap of unknown length
* 8978 10256: contig of 1279 bp in length
* 10257 10356: gap of unknown length
* 10357 11356: contig of 1000 bp in length
* 11357 11456: gap of unknown length
* 11457 12470: contig of 1014 bp in length
* 12471 12571: gap of unknown length
* 12571 13862: contig of 1291 bp in length
* 13862 13961: gap of unknown length
* 13961 15061: contig of 1099 bp in length
* 15061 15160: gap of unknown length
* 15160 16252: contig of 1092 bp in length
* 16253 16352: gap of unknown length
* 16353 17470: contig of 1118 bp in length
* 17471 17570: gap of unknown length
* 17571 18837: contig of 1267 bp in length
* 18838 20247: gap of unknown length
* 20247 20347: contig of 1309 bp in length
* 20347 21516: gap of unknown length
* 21516 21615: contig of 1169 bp in length
* 21615 22752: contig of 1137 bp in length
* 22753 24438: contig of 1586 bp in length
* 24439 24538: gap of unknown length
* 24539 25615: contig of 1077 bp in length
* 25616 25715: gap of unknown length
* 25716 26919: contig of 1203 bp in length
* 26919 27019: gap of unknown length
* 27019 28161: contig of 1143 bp in length
* 28162 28261: gap of unknown length
* 28262 29906: contig of 1645 bp in length
* 29907 30006: gap of unknown length
* 30007 31335: contig of 1329 bp in length
* 31336 31436: gap of unknown length
* 31436 32662: contig of 1226 bp in length
* 32662 32761: gap of unknown length
* 32761 34093: contig of 1332 bp in length
* 34093 34194: gap of unknown length
* 34194 35482: contig of 1289 bp in length
* 35483 35582: gap of unknown length
* 35583 37001: contig of 1419 bp in length
* 37002 38337: contig of 1236 bp in length
* 38337 38437: gap of unknown length
* 38438 39825: contig of 1388 bp in length
* 39825 41162: gap of unknown length
* 41162 41262: contig of 1237 bp in length
* 41263 42872: contig of 1610 bp in length
* 42873 42972: gap of unknown length
* 42973 44118: contig of 1146 bp in length
* 44119 44219: gap of unknown length
* 44219 45913: contig of 1699 bp in length
* 45913 46013: gap of unknown length
* 46014 47295: contig of 1282 bp in length
* 47296 47395: gap of unknown length
* 47396 48501: gap of unknown length
* 48501 48600: gap of unknown length
* 48601 50670: contig of 1105 bp in length
* 50671 50770: gap of unknown length
* 50771 52744: contig of 2070 bp in length
* 52745 52844: gap of unknown length
* 52845 54280: contig of 1974 bp in length
* 54281 54380: gap of unknown length
* 54381 56311: gap of unknown length
* 56312 56411: contig of 1931 bp in length
* 56412 59166: gap of unknown length
* 59167 59266: contig of 2755 bp in length
* 59267 62681: gap of unknown length
* 62682 62781: contig of 3415 bp in length
* 62782 65691: gap of unknown length
* 65691: contig of 2910 bp in length

AL627280	245050 bp	DNA	linear	BCF 06-JUN-2003
LOCUS				
DEFINITION	<i>Salmonella enterica</i> serovar Typh (Salmonella typhi) strain CT18,			
ACCESSION	complete chromosome; segment 16/20.			
VERSION	AL627280.AL513382			
KEYWORDS				
SOURCE	AL627280.1 GI:16504729			
ORGANISM				
REFERENCE	Salmonella enterica subsp. enterica serovar Typhl.			
AUTHORS	Salmonella enterica subsp. enterica serovar Typhl Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella.			
1 (bases 1 to 245050)				
Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D., Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G., Sebaiha,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T., Comerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.S., Farar,J., Felwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S., Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moute,S., O'Geare,P., Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J., Stevens,K., Whitehead,S. and Barrall,B.G.				
Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhl CT18				
Nature 413 (6858), 848-852 (2001)				
JOURNAL				
MEDLINE	21534947			
PUBMED	11677608			
REFERENCE	2 (bases 1 to 245050)			
AUTHORS	Parkhill,J.			
TITLE	Direct Submission			
JOURNAL	Submitted (25-OCT-2001) Submitted on behalf of the Salmonella sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK E-mail: parkhill@sanger.ac.uk			
COMMENT	Notes: Details of S. typhi sequencing at the Sanger Centre are available on the World Wide Web. (URL, http://www.sanger.ac.uk/projects/S-typhi/). Location/Qualifiers 1..245050 /organism="Salmonella enterica subsp. enterica serovar Typhl"			
FEATURES				
SOURCE				

FEATURES	source
Localblast	1. .245050
	/organism="salmonella enterica subsp. enterica serovar typhl"
	/strain="Cr18"
	/db_xref="taxon:90370"
gene	319. .1728
	/gene="STY3874"
	/note="glna"
CDS	319. .1728
	/gene="STY3874"
	/note="Orthologue of E. coli glnA (GlnA-ECOLI); Fasta hit to GlnA-ECOLI (468 aa), 98% identity in 468 aa overlap"

```

/cease=1
/transl_table=1
/product="glutamine synthetase"
/protein_id="CAD03093.1"

```

```

/misc_feature
/misc_feature
misc_feature

/ab_xref="SWISS-PROT.i06201
/translation="MSAEHLVLMLEHNEKRVDFLPNDSTKIDLELIPAHQVNAEEF
EKKMKDGGSIGKMGKIGNSDVLVLPDSTVIDEPDFADSLITKIDLELIPAHQVNAEEF
RQDRSLAKRAEDYLRATGIGIADVLVLPDSTVIDEPDFADSLITKIDLELIPAHQVNAEEF
SRKYEGCKNGKGRPGKGGYEPVVDASODI RSMCKLTMQONKGLYVEAHNDEHATAG
QNEVATRPDMTKKADEIDYKVVVHNVAHNRGKGTATYTPKPKNGDSGMCNIMSLA
KNGTINLPSDGTAGLSEALYIGCVIKAKAINLALNAPYLCFAALLMGLOGIKNKIRGGA
YSANRKSASTRIPIVNASPKARILEVFDPAPAINPYLCFAALLMGLOGIKNKIRGGA
MDKRLDLPPEPAKLTPOVAGLSEALINLDDREFLKAGVFDDEALIDATIALRREE
DDVRMTPTHPVEFELYTSV"
364..1434
/gene="STY3874"
/note="Pfam match to entry PF00120 gln-synt, Glutamine
synthetase, score 838.00, E-value 2.4e-266"
466..522
/gene="STY3874"
/note="PS00180 Glutamine synthetase signature 1"
1093..1140
/gene="STY3874"

```

```

/misc-feature /note="PS00181 Glutamine synthetase putative ATP-binding
region signature"
/gene="STY3874"
/site="PS00182 Glutamine synthetase class-I adenylation
site"
gene 2003.3052
/gene="STY3875"
/note="glnL"
CDS 2003.3052
/gene="STY3875"
/note="Orthologue of E. coli glnL (NTRB_ECOLI): Fasta hit
to NTRB_ECOLI (349 aa), 93% identity in 349 aa overlap"
/transl_table=11
/product="Two-component system sensory histidine kinase"
/protein_id="CAD03094.1"
/db_xref="GI:16504731"
/db_xref="SWISS-PROT:P41788"
/translation="MAGIOPDAGQILNLSIVLVDDALAHYANPAQAOLLAQSS
RIKLEPELLELSYSLNIDLRSLAAGCGFTNEVTLVDSKSLSLTAORLPD
FILEMAPMDNORLSOLOAOIARDLVRLGLAHEIKNPGLGLGAOLSKALP
RLIRYDPSLEPEIIOVADRLNLDLGFHGHMHTESIKHVAERVALYSMLPBNV
YLAARIDEDNCPPIPHLDLTFYPMVSGREGTGLSLIARNLIDHACKIEETS
MGEHFEFVYLPKIK"
/misc-feature 2417.3040
/gene="STY3875"
/note="Pfam match to entry PF00512 signal, Histidine
kinase, score 244.40, E-value 1.6e-69"
CDS 3061.4470
/gene="STY3876"
/note="glnG"
3061.4470
/gene="STY3876"
/note="Fasta hit to HYDG_ECOLI (441 aa), 44% identity in
462 aa overlap"
Fasta hit to YGA_ECOLI (529 aa), 36% identity in 451 aa
overlap
Fasta hit to ATQC_ECOLI (461 aa), 43% identity in 468 aa
overlap
Fasta hit to YFHA_ECOLI (444 aa), 44% identity in 381 aa
overlap
Orthologue of E. coli glnG (NTRC_ECOLI): Fasta hit to
NTRC_ECOLI (469 aa), 96% identity in 469 aa overlap"
/transl_table=11
/product="Two-component system, response regulator"
/protein_id="CAD03095.1"
/db_xref="GI:16504732"
/db_xref="SPTRMBL:082233"
/translation="MORGIVWVDDSSIRWLERALAGAGLCTFENGNEVLAALA
SKTPDVLSDIRMGMDGLALLKQIKORHPMLVITMTAHSIDLDAVSAQCAFYL
PKPDIDEAVAIYBRAISHIOEQOQPRNTEVNGPTDMGEAPMODFRTIGRSR
SIVLINGSCTGKELVYALHHRSPRAKPFIALNMAIPKDLIESTLFEHKGAT
GANTIRGREGQADGGTLFDELIGDMLDVOTRLRLVADOCFYVGGIAYKDVAVI
IAATHONLEKRVQEGFREDLFPHLNVIRIHLPPREREDIPRLARHFLQVAREIG
VEAKLHPETETALTRIANPQVROLENTCEMLTYMAAGQEVLTQDLPFGLFESAPD
SPSHLPDPSWATLILAQMDARALRSQNLSDAPELERTLLTALRHHTGHOEAKK
LIGWGRNLTTRKELGME"
/misc-feature 3070.3405
/gene="STY3876"
/note="Pfam match to entry PF00072 response_reg, Response
regulator receiver domain, score 141.50, E-value 1.5e-38"
/misc-feature 3478.4173
/gene="STY3876"
/transcription factors, score 515.80, E-value 3.1e-151"
/misc-feature 3550.3591
/gene="STY3876"
/note="PS00675 Sigma-54 interaction domain ATP-binding
region A signature"
/misc-feature 3736.3783

```

```

/misc-feature /note="STY3876"
/gene="PS00676 Sigma-54 interaction domain ATP-binding
region B signature"
/misc-feature 4117.4146
/gene="STY3876"
/note="PS00688 Sigma-54 interaction domain C-terminal part
signature"
gene complement(4783.6156)
/gene="STY3877"
/note="hemN"
CDS complement(4783.6156)
/gene="STY3877"
/note="Orthologue of E. coli hemN (HEMN_ECOLI): Fasta hit
to HEMN_ECOLI (457 aa), 93% identity in 457 aa overlap"
/transl_table=11
/product="oxygen-independent coproporphyrinogen III
oxidase"
/protein_id="CAD03096.1"
/db_xref="GI:16504733"
/db_xref="SWISS-PROT:P37129"
/translation="MSEQIDMDLALIQKYNVSGPRYTPYLAESSEDEDAFLQA
VARYPERPLSLYVHIPEFCHKLCYFCGCKKIVTROKHADQYDLAEQELRHAPLAD
RHYVSLHMGSGTPYVLYNKAQISRLMTLRENFHNTDAEISIEVDREILDVHLR
AEGFNRLSMGVODPFNKVORLYNREODEEFTFALLNNAIDIGFTNIDIGLPOR
PESFAVTLKRVTELNDRLSYFNALHDEFLPAORAKTKDADLSAQQKIDIIQETYS
LTQAGYOFIGMDHFAKRPDELAIVAKREGVLRHNFQGTGGDGLGKGVSAISMG
GYMONOKELRYKVOYDERGNALWRGITLIRDCDIRDYIKALICNFRIDPNAVDQW
GLHFAEYREDQLLPLAKGDIVDISEKIQVYAKGRLLIRNICMCFAYLRQAKM
OQSFYVI"
gene complement(6345.6860)
/gene="STY3878"
/note="yihI"
CDS complement(6345.6860)
/gene="STY3878"
/note="Orthologue of E. coli yihI (YIH_ECOLI): Fasta hit
to YIH_ECOLI (169 aa), 85% identity in 169 aa overlap.
Note this protein is extremely hydrophobic"
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD03097.1"
/db_xref="GI:16504734"
/db_xref="SWISS-PROT:P37130"
/translation="MKKPSAPRSKAFKQRRKREELNDEARDRLKKRHGARGS
RAAGNSAGSGNONOQDPRIGSKTIPVLEGTETKYOQKPKSEKPMISPAEIDL
EDMERLDALLELENGETLSADDAQWDAKLRLIDELMOKLGLSYDDDEDEDEBKO
EDMRRLRGNN"
gene complement(6979.7203)
/gene="STY3879"
complement(6979.7203)
/note="no significant database hits. Note this protein is
extremely hydrophobic"

```

```

Query Match 61.5% Score 184.4; DB 1; Length 245050;
Best Local Similarity 81.9%; Pred. No. 5.7e-49;
Matches 249; Conservative 0; Mismatches 51; Indels 4; Gaps 3;
QY 1 ATTCAATCTGTTGATCGTGGTGTGGCTGATGATTAAGCATCCCTGCTGAATAA 60
DB 110092 ATTCATCTGTTGTTGTCGTTATTTGCATGATAGTTAAGGATCCGTTGA 110151
QY 61 ACATCATATTACGTGGCAGCTGCGGCGCTATGCGACTTAAAGTTGCGTGC-CC 119
DB 110152 ACATCACAGAGTGGCAGCTGCGGCGGTAAAGCAGCTTAACATAGCTCTCCAA 110211
QY 120 TCAATCATCAATAGACATAAAGTCAAAAGTCCGCTGATAGAGCTTGCAAAAGT 179
DB 110212 CCACTCATCAATAGACATAAAGTCAAAAGTCCGCTGATAGAGCTTGCAAAAGT 110271
QY 180 TCAATTCAGACCACTTTTAAATCGTAGCCGAT--GAGACGCGCCTGATGGGTGTCT 237

```



```

DB 110331 ACCTATTGACACCATTTATAGACGTAGTGAAGGCGACCCGATGGGTCT 110331
QY 238 GGCTA-CCTGACCTGTCATGTCGAGAGTCTTATCTGCTGATTCAGAGCTAAT 296
DB 110332 GGCAATCCGCGCTTCATGATGAGAGTCTGACATTCCTGCTATTCAGAGTTAT 110391
QY 297 GATT 300
DB 110392 GATT 110395

RESULT 12
AE008877/c 22492 bp DNA linear BCT 31-JUL-2002
DEFINITION Salmoneella typhimurium LT2, section 181 of 220 of the complete
ACCESSION AE008877 AE006468
VERSION AE008877.1 GI:16422367
KEYWORDS
SOURCE
ORGANISM Salmoneella typhimurium LT2.
REFERENCE 1 (bases 1 to 22492)
AUTHORS Mclelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W.,
Latreille, P., Courtney, L., Portolillo, S., All, J., Dante, M., Du, F.,
Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A.,
Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, L., Miller, W.,
Stonking, T., Nhan, M., Waterston, R. and Wilson, R.K.
TITLE Complete genome sequence of Salmoneella enterica serovar Typhimurium
LT2
JOURNAL Nature 413 (6858), 852-856 (2001)
MEDLINE 21354948
PUBMED 11677609
REFERENCE 2 (bases 1 to 22492)
AUTHORS The Salmoneella typhimurium Genome Sequencing Project.
TITLE Direct Submission
JOURNAL Submitted (29-MAR-2001) Genome Sequencing Center, Department of
Genetics, Washington University School of Medicine, 4444 Forest
Park Boulevard, St. Louis, MO 63108, USA
COMMENT Supported by NIH Grant 5U 01 AI43283

COMMENT
Coding sequences below are predicted from manually evaluated
computer analysis, using similarity information and the programs;
GLIMMER, http://www.tigr.org/softlab/glimmer/glimmer.html and
GeneMark; http://opal.biology.gatech.edu/GeneMark/
EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
and Pedro Romero and Peter Karp at EcoCyc;
http://ecocyc.org/ecocyc/

The analyses of ribosome binding sites and promoter binding sites
were kindly provided by Heladia Salgado, Julio Collado-VIDES and
Reguondb;
http://kinich.cifn.unam.mx:8850/db/reguondb\_intro.frameset

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistries or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one m33 subclone.

FEATURES
Source
1..22492
/organism="Salmoneella typhimurium LT2"
/strain="LT2; SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
/note="LT2"
complement(78..278)
/gene="lydL"

CDS
10_signal
/note="putative RBS for lydL; Reguondb:STMS1H003766"
complement(254..262)
/gene="lydL"
/note="putative -10_signal for lydL;
Reguondb:STMTLH004613"
complement(264..292)
/note="putative binding site for CRP, Reguondb:
STMS1H000038"
/bound_molety="CRP"
complement(270..278)
/gene="lydL"
/note="putative -35_signal for lydL;
Reguondb:STMTLH004613"
join(699..790,789..813)
/gene="ysdA"
/note="ribosomal slippage"
/codon_start=1
/transl_table=11
/product="ysdA"
/protein_id="AAM89474.1"
/db_xref="GI:22025623"
/translation="MSSTRFLAKGIACAPAGNRRKQTSAGSAPLNHRRV"
810..899
/gene="ysdB"
/note="STM3796B"
810..899
/gene="ysdB"
/codon_start=1
/transl_table=11
/product="conserved protein in the LexA regulon"
/protein_id="AAL22369.1"
/db_xref="GI:16422369"
/translation="MSVVDTITLIKILVALQLDAVLKYLK"
92..1852
/gene="STM3796A"
92..1852
/gene="STM3796A"
/codon_start=1
/transl_table=11
/product="integral membrane protein"
/protein_id="AAL22657.1"
/db_xref="GI:16422370"
/translation="MTSPTPDAMTISVFCILFPAALLHASMNAIVKAGNDKLTAAIG
VSGSAVAAVAILIPSPPOPAHSIPDAISALGVYTVLVAKTQYSDMSQTYPLMR
GTAPLVAIVSVFLGDSLSLAWGVIVICMALLIGMAGNRRASQRYVALTNACF
IAGYTLVGTGRTSETALVGTLSFUNGACILTMWAIHARRASRIADQMKKGF
AGIGTMSYGLALNMTQAPLAVVAALRETSILFGALILWLLKEVAGLRLVAAGGI
ALGAILRLS"
2055..3239
/gene="emrD"
/note="STM3798"
2055..3239
/gene="emrD"
/note="similar to E. coli 2-module integral membrane pump;
multidrug resistance (AAC76696.1); Blastp hit to
AAC76696.1 (396 aa), 92% identity in aa 3 - 396"

```

erv Match

SECRET

500

1
2
3
4
5
6
7

SUMMARY

ORGANIZATIONAL

NEW ENGLAND

TITTLE

REFERENCES

WOLFF

AULHOK,

Page 2

TITLE
JOURNAL

COMMENT

Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamali, A., Karatas, A., Kells, C., Landers, P., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Strange-Thomann, N., Stojanovic, N., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 21, 2002 this sequence version replaced gi:17194989.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu

Project Information
Center project name: L18603
Center clone name: 336_E-5

Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 17783 bases at least Q40
Consensus quality: 17985 bases at least Q30
Consensus quality: 180508 bases at least Q20
Insert size: 163000; agarose-ftp
Insert size: 180877; sum-of-coverage
Quality coverage: 7.3 in Q20 bases; sum-of-coverage
Quality coverage: 6.6 in Q20 bases; sum-of-coverage

NOTE: This is a 'working draft' sequence. It currently consists of 12 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

933 1032: contig of 932 bp in length
1033 1556: contig of 524 bp in length
1557 1656: gap of 100 bp
1657 2569: contig of 913 bp in length
2570 2669: gap of 100 bp
2670 4587: contig of 1918 bp in length
4588 4687: gap of 100 bp
4688 6879: contig of 2192 bp in length
6880 6979: gap of 100 bp
6980 10112: contig of 3133 bp in length
10113 10212: gap of 100 bp
10213 14936: contig of 4724 bp in length
14937 15036: gap of 100 bp
15037 27001: contig of 11965 bp in length
27002 27101: gap of 100 bp
27102 114046: contig of 86495 bp in length
114047 114146: gap of 100 bp
114147 131634: contig of 17488 bp in length
131635 131734: gap of 100 bp
131735 167389: contig of 35655 bp in length
167390 167489: gap of 100 bp
167490 181977: contig of 14488 bp in length.
Location/Qualifiers
1. 181977

FEATURES
source

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-336E5"
/clone_11p="RC1-23 Female Mouse BAC"
1. 932
/note="assembly_fragment
vector_side:left"
1033. 1556
/note="assembly_fragment"
1657. 2569
/note="assembly_fragment"
2670. 4587
/note="assembly_fragment"
4688. 6879
/note="assembly_fragment"
6980. 10112
/note="assembly_fragment"
10213. 14936
/note="assembly_fragment"
15037. 27001
/note="assembly_fragment"
27102. 114046
/note="assembly_fragment"
114147. 131634
/note="assembly_fragment"
131735. 167389
/note="assembly_fragment"
167490. 181977
/note="assembly_fragment
vector_side:right"

BASE COUNT 53329 a 35929 c 35700 g 55913 t 1106 others
ORIGIN

Query Match 12.3%; Score 36.8; DB 2; Length 181977;
Best Local Similarity 51.9%; Pred. No. 1.4;
Matches 107; Conservative 0; Mismatches 97; Indels 2; Gaps 1;

QY 55 AAATAATCATCATCATCTAGCTCCGACCTGCGGCTATCGCACTTTAACTTCTGTCG 114
DB 10984 AAATACATTAATAATACAAAGTATTTCACCGGCTATGACAC--CACCAATTGGTCATG 10927
QY 115 CCCCCAGCTGATCATCATGACCATTAACCTGCAAAAAAAGTCCGCTGATAAGCTTGA 174
DB 10926 AGAATATGCTGTCGCTTGACATTTCTGTGTAACAACAGCTGCTTCACTCATTTGG 10867
QY 175 AAAGTTCATTTCAGACCCATTTTACATCTGACCCATGAGACGCGCTGATGGGT 234
DB 10866 AAGCTGCTTCCCGCCACCTGGGATTTTATCCCACTGAGGAGTGCGTATTTCT 10807
QY 235 TCTGCTACCTGACCTGTCATTTGG 260
DB 10806 ACTGCAATTTGCTATACCAACAGTG 10781

RESULT 14

AP005103/c 160174 bp DNA linear HTG 20-APR-2002

LOCUS Oryza sativa (japonica cultivar-group) chromosome 7 clone

DEFINITION OSJNBa0036M16, *** SEQUENCING IN PROGRESS ***; in ordered pieces.

ACCESSION AP005103.1 GI:20218991

VERSION HTG: PHASE2.

KEYWORDS Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA, clone:OSJNBa0036M16

ORGANISM Oryza sativa (japonica cultivar-group); Embryophyta; Tracheophyta; Eukaryota; Viridiplantae; Streptophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzae; Oryza.

REFERENCE 1 Sasaki, T., Matsumoto, T. and Katayose, Y.
Oryza sativa nipponbare(GAS) genomic DNA, chromosome 7, BAC

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

clone: OSJNBa0036M16
Published Only in Database (2002)
2 (bases 1 to 160174)

COMMENT

FEATURES

- * The contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submitter. This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.
- * NOTE: This is a 'working draft' sequence.
- * This sequence will be replaced
- * by the finished sequence as soon as it is available and the accession number will be preserved.
- * location/contig will be preserved.

[illegible]

Db 120337 TTGTAGACCTAGCTT 120323

RESULT 15	
CNS06C81/c	DNA
LOCUS	195007 bp
DEFINITION	linear
ACCESSION	PRI 28-APR-2001
VERSION	Human chromosome 14 DNA sequence BAC R-233L14 of library RPCT-11
KEYWORDS	AL391749 complete sequence.
SOURCE	AL391749.4 GI:13697305
	HGT.
	human.

JOURNAL
COMMENT

Submitted (27-Apr-2001) Genoscope - Centre National de Séquençage
BP 191 91006 Evry cedex - FRANCE (E-mail : segeir@genoscope.cns.fr)
- Web : www.genoscope.cns.fr
On April 30, 2001 this sequence version replaced gi:12619149.

Genome Center

STS 176049. .176250
/note="matching EMBL:T89997

RHdb:RH26726

dbSTS:STS19237

Identified using the e-PCR software (G. Schuler)"

STS 187792. .187892
/note="matching EMBL:G20753

RHdb:RH17210

dbSTS:STS20239

Identified using the e-PCR software (G. Schuler)"

BASE COUNT 59443 a 40899 c 40806 g 53857 t 2 others

ORIGIN

Query Match 12.1%; Score 36.2; DB 9; Length 195007;

Best local Similarity 57.5%; Pred. No. 2.2; Mismatches 48; Indels 0; Gaps 0;

Matches 65; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

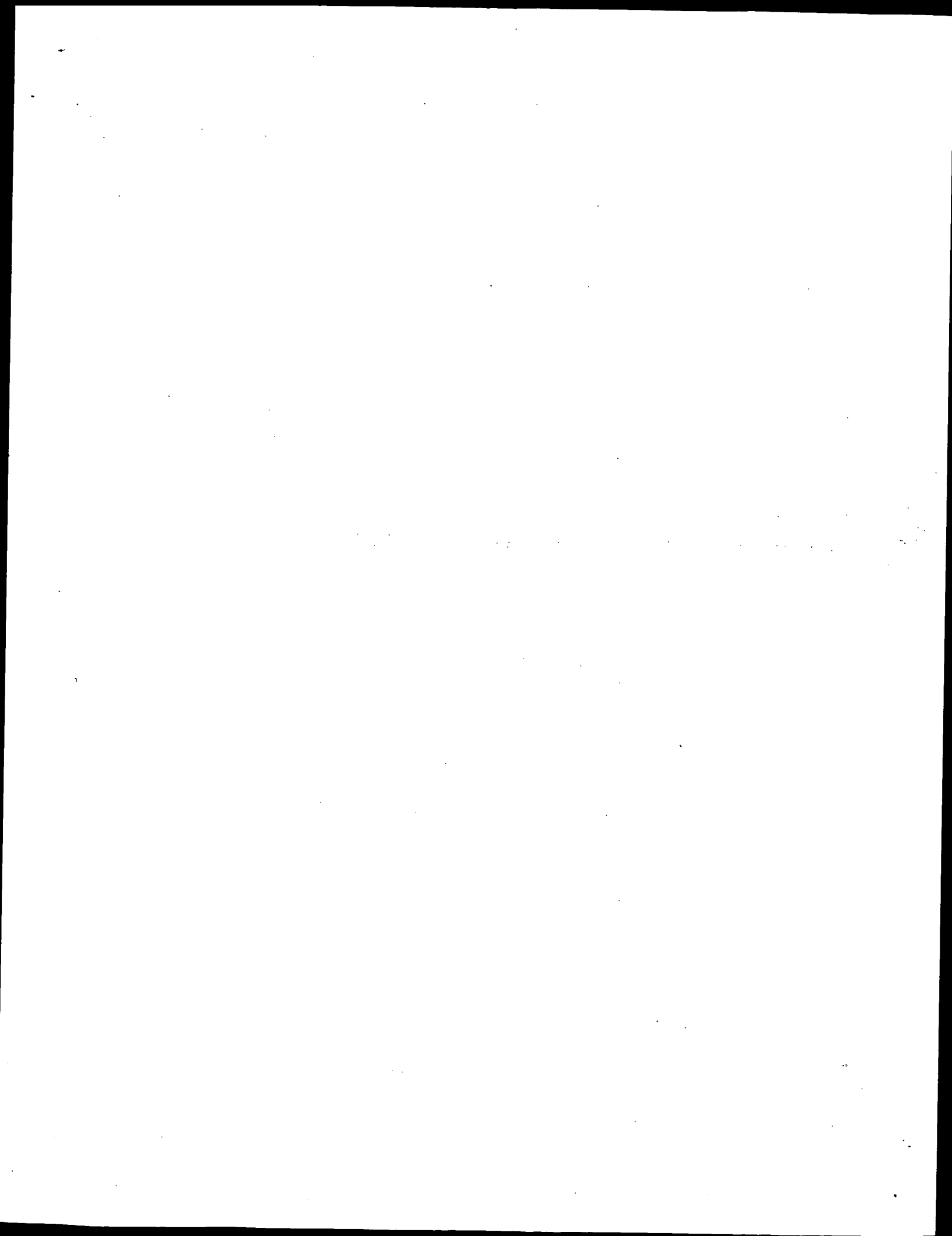
QY 106 TTGCTGCTGCCCCCTCAGTCTATGCAATGACCATTAACGCAAAAAAGTCCGCTGAT 165

Db 138137 TTGTCTCAGGGAGGAGATTGAGATGACATGACAAACGCAAAAAATTAACCCACCT 138078

QY 166 AAGGCTTGAAAGTCAATTCAGACCCATTTTACATCGTAGCCGATGAGGA 218

Db 138077 GTTCTTGAGATTTAATTTCAACAGACTTTTACATTTCCACCTGGGAGGA 138025

Search completed: May 5, 2003, 19:21:18
Job time: 2053.7 secs



PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

Claim 1; SEQ ID NO 29473; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (I) and its binding partners are useful in medical disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations in responsible for genetic disorders or other traits to assess biodiversity, and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

Sequence 815 BP; 210 A; 203 C; 187 G; 215 T; 0 other;

Query Match
Best Local Similarity 78.8%; Score 236.4; DB 23; Length 815;
Matches 240; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

YY 55 AAAAATCATCATCATTCAGTCGACATGTCGGCGCTATCCGACATTACGTTTCGTCG 114
DB 8 AAAAATCATCATCATTCAGTCGACATGTCGGCGCTATCCGACATTACGTTTCGTCG 67
YY 115 CCCCCTAGTCTATGCAATGACCATTAACAAAAAGTCCGCTGTAAGGCTTGA 174
DB 68 CCCCCTAGTCTATGCAATGACCATTAACAAAAAGTCCGCTGTAAGGCTTGA 127
YY 175 AAGTTCATTCAGACACATTTTTCATGCTAGCCGATAGAGACCGCTGATGGGTGT 234
DB 128 AAGTTCATTCAGACACATTTTTCATGCTAGCCGATAGAGACCGCTGATGGGTGT 187
YY 235 TCGGCTACCTGACCTGTCATTTGGAAGGCTTACATTCGCTGATTTGAGAGCTA 294
DB 188 TCGGCTACCTGACCTGTCATTTGGAAGGCTTACATTCGCTGATTTGAGAGCTA 247
YY 295 TTGATT 300
DB 248 TTGATT 253

RESULT 2
AAH03353
ID AAH03353 standard; cDNA; 866 BP.

AC AAH03353;
XX
XX 26-JUN-2001 (first entry)
DE Human cDNA clone (5'-primer) SEQ ID NO:188.

OS Homo sapiens.
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX EPI074617-A2.
XX
XX 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.

Claim 1; SEQ ID 188; 2537p + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the of an oligonucleotide comprises a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human amino acid sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 866 BP; 219 A; 228 C; 215 G; 201 T; 3 other;

Query Match
Best Local Similarity 11.5%; Score 34.4; DB 22; Length 866;
Matches 71; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

YY 153 AAGTCCGCTGATTAAGGCTTGAAGGCTTATTCAGACCCATTTTACATTCGATGCGGA 212
DB 353 AAGGCTGCTGTTGCTTACAGAGTCATTTATTCACACATTAAGACATGAAGC 412
YY 213 TGAGAGCGCGCTGATGAGGCTTTCGCTACCTGCTGATGGAAGCTTACA 272
DB 413 TGTGAGACCTCTCTGATATTCCTGCAACCAACGAGCTATGCTCTATATCA 472
YY 273 TTTGCTGATT 284
DB 473 ACCATTCCAATT 484

RESULT 3
AAC8997
ID AAC8997 standard; cDNA; 1791 BP.

AC AAC8997;
XX
XX 09-MAR-2001 (first entry)
DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:225.

XX Human; pancreas; pancreatic cancer; antigen;
 KM detection; diagnosis; identification; cytostatic; neuroprotective;
 KM neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological
 KM antineoplastic; cardiac; gene therapy; chromosome mapping
 KM linkage analysis; tissue identification; tissue typing; forensic;
 KM neural; immune system; muscular; reproductive; gastrointestinal;
 KM pulmonary; cardiovascular; renal; proliferative; ss.

OS Homo sapiens

PN WO200055320-A1.

PD 21-SEP-2000.

08-MAR-2000; 2000WO-US05989

PR 12-MAR-1999; 99US-01242/0.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM,

DR WPI; 2000-579444/54

DR P-PSDB; AAB54232.

XX PT New nucleic acid that is a pancreatic cancer antigen for preventing,
XX PT treating, or ameliorating a medical condition, particular pancreatic
XX PT cancer, or for use in assays for diagnosing a pathological condition
PS Claim 1; Page 668-669; 1379pp; English.

XX AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC proteins, called pancreatic cancer antigens, given in AAB54408 to
CC AAB54466. The human pancreatic cancer antigens have cytostatic,
CC neuroprotective, cardioprotic, immunomodulatory, relaxant, contraceptive,
CC gynaecological, radiant and antiinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for
CC preventing, treating, or ameliorating a medical condition or in assays
CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to
CC detect, treat or prevent pancreatic disorders, especially cancer.
CC Agonists and antagonists to the antigens can be screened for. The
CC pancreatic cancer antigen polynucleotides can be used to design nucleic
CC acid hybridisation probes that can be used in chromosome mapping, linkage
CC analysis, tissue identification and/or typing and a variety of forensic
CC and diagnostic methods. The proteins can be used to generate antibodies
CC which are used to purify, detect and target the polypeptides. The
CC both in vivo and in vitro diagnostic and therapeutic methods. The
CC proteins can be used to treat or prevent neural, immune system, muscular
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.

aa Sequence 1791 BP; 497 A; 422 C; 420 G; 452 T; 0 other
sq

Query Match	11.5%	Score 34.4;	DB 21;	Length 1791;
Best Local Similarity	53.8%	Pred. No. 0.19;		
Matches	71;	Conservative	0;	Mismatches 61;
				Indels 0;
				Gaps 0

153 AAAGTCGCTGATAAGCCTTGAAAGTTCATTTCCAGACCATTTTACATCGTAGCCGA 212

Db 276 A A A G C T G C T G G T T T G C C T A G A A G A G T C A T T T A T T A T T C A C A C A C A T T T A A G A C A T G A A G C 335

213 TGAGGACGGCCTGATGGGTGTTCTGGCTACCTGACCTGTCATTGTGGAAGTCTTACA 272

Db 336 TGTGAGAGACCTCTGGATATTCCTGCAACCCACAGGTCTATGTCTCTCTATCA 395

273 TTCTCGCTGATT 284

Db 396 ACCATTCCAATT 407

RESULT 4
AAH13701
ID AAH13701 standard; cDNA; 1853 BP.

AC ААН13701

DT 26-JUN-2001 (first entry)

Human CDNA sequence SEQ ID NO:10578

KW Human; primer; detection; diagnosis; antisense therapy; gene expression

OS Homo sapiens.

PN EP1074617-A2

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126

PR 29-JUL-1999; 99JP-0248036

PR 11-JAN-2000; 2000JP-0118776

PR 09-JUN-2000; 2000JP-0241899

PA (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa M
PI Ota T, Isogai T, Nishikawa M

XX
2001 310740/3A

XX
for synthesizing polynucleotides, particularly the 560

PT full-length cDNAs defined in the specification, and for the identification of the abnormality of the proteins encoded by the

PT full-length cDNAs -

PS Claim 8; SEQ ID 105/6; 253/pp + CD ROM, English

AA The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification, where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises at least 15 nucleotides and the combination of
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primers are useful for synthesizing polynucleotides
CC in gene therapy. The primers are useful for synthesizing polynucleotides
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH93893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX	Sequence	1853 BP;	503 A;	448 C;	445 G;	457 T;	0 other
SQ							

Query Match	11.5%;	Score 34.4;	DB 22;	Length 1853;
Best Local Similarity	53.8%;	Pred. No. 0.19;		
Matches	71;	Conservative	0;	Mismatches 61;
				Indels 0;
				Gaps 0;

QY 153 AAAGTCCGCTGATTAAGCTTGAAGAATTCATTTCCAGACCCATTTTACATCGTAGCCGA 212

OY 213 TGAGGAGCGGCGCTGATGGTGTCTGCTACCGACCTTCATTTGAGAGCTTTACA 272
 DB 413 TGTGTAAGACCCCTCGATATTCCTCGCAACCAACAGCTGTATGTCTCTCTATCA 472
 OY 273 TTCTCGCTATT 284
 DB 473 ACCATTCCATT 484
 RESULT 5
 AAC76764
 ID AAC76764 standard; cDNA: 2162 BP.
 AC AAC76764;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORF2319 polynucleotide sequence SEQ ID NO:4637.
 XX
 KW Human; open reading frame; ORF; detection; cytostatic; hepatotropic;
 KW vulnary; antiparasitic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW antiviral; antibacterial; antifungal; immunosuppressive; antineoplastic;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypochromidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antineoplastic disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0340763.
 XX
 PA (CUBA-) CURAGEN CORP.
 XX
 PI Shimkels RA, Leach M;
 XX
 DR WPI: 2000-602362/57.
 DR P-PSDB; ABA42555.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 5; Page 3835-3836; 5507pp; English.
 XX
 AC AAC74446 to AAC77606 encode the proteins given in ABA40237 to ABA43397,
 CC which represent the human ORF open reading frames 1 to 3161. The ORF
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antiparasitic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antidiabetic; immunosuppressant;
 CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antineoplastic; antibacterial; antiviral; antifungal; antineoplastic;
 CC antihypertensive; antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORF-associated disorder. The
 CC nucleic acids can be used to express ORF proteins in gene therapy.

CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypochromidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antineoplastic disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 SO Sequence 2162 BP; 575 A; 511 C; 526 G; 548 T; 2 other:
 Query Match 11.5%; Score 34.4; DB 21; Length 2162;
 Best Local Similarity 53.8%; Pred. No. 0.21;
 Matches 71; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
 OY 153 AAGTCCGCTGATTAAGCTTGAAGTTCATTTCCAGACCAATTTTACATGTCGAGCGA 212
 DB 401 AAGGCTGCTGCTTGTTCCTGAGAGACTCATTTATTCACAACTTAAGACATGAAGC 460
 OY 213 TGAGGAGCGGCGCTGATGGTGTCTGCTACCGACCTTCATTTGAGAGCTTTACA 272
 DB 461 TGTGTAAGACCCCTCGATATTCCTCGCAACCAACAGCTGTATGTCTCTCTATCA 520
 OY 273 TTCTCGCTATT 284
 DB 521 ACCATTCCATT 532
 RESULT 6
 ABL89998
 ID ABL89998 standard; cDNA: 2318 BP.
 XX
 AC ABL89998;
 XX
 DT 24-MAY-2002 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 560.
 XX
 KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antineoplastic; antiparasitic;
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200190304-A2.
 XX
 PD 29-NOV-2001.
 XX
 PF 18-MAY-2001; 2001WO-US16450.
 XX
 PR 19-MAY-2000; 2000US-205515P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Birse CE, Rosen CA;
 XX
 DR WPI: 2002-122018/16.
 DR P-PSDB; ABB89589.
 XX
 PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
 PT prevention of neural, immune system, muscular, reproductive,
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
 PT disorders -
 XX
 PS Claim 4; SEQ ID NO 560; 2081pp + Sequence Listing; English.
 XX
 AC The invention relates to novel genes (ABL89449-ABL90853) and proteins
 CC (ABB89440-ABB90444), useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.

CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia and
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pat_sequences.

XX Sequence 2318 BP: 629 A; 538 C; 569 G; 582 T; 0 other;

Query Match 11.5%; Score 34.4; DB 24; Length 2318;
Best Local Similarity 53.8%; Pred. No. 0.22; Mismatches 0; Gaps 0;
Matches 71; Conservative 0; Indels 0;

OY 153 AAGTCCGCTGATAGGCTGAAAGTTCATTTCCAGACCCATTTTACATCGTAGCCGA 212
DB 436 AAGGCTGCTGTTGCTTCCAGAGAGTCCATTTATTCACAAATTAAGACATGAAGC 495
OY 213 TGAGAGCGCGCTGATGGCTGTGCTGCTGACCTGCTGCTGCTGAGAGCTTACA 272
DB 496 TGTGAGAGCGCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTA 555
OY 273 TGTGCTGCTGAT 284
DB 556 ACCATTCGAAT 567

RESULT 7

AA121446/C
ID AA121446 standard; cDNA: 720 BP.

XX AA121446;

DT 07-DEC-2001 (first entry)

DE Human breast cancer expressed polynucleotide 13903.

KW Human; breast cancer; cell marker; cytostatic; ss.

OS Homo sapiens.

PN WO200151628-A2.

PD 19-JUL-2001.

PF 10-JAN-2001; 2001WO-US00798.

PR 14-JAN-2000; 2000US-0176077.

PR 14-MAR-2000; 2000US-0189167.

PR 24-MAR-2000; 2000US-0192099.

PR 29-MAR-2000; 2000US-0193480.

PR 15-MAY-2000; 2000US-0205230.

PR 09-JUN-2000; 2000US-0211315.

PR 25-JUL-2000; 2000US-0220534.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Lillie J, Xu Y, Wang Y, Steinmann K.

XX WPI: 2001-451856/48.

XX New peptide useful as a marker for the diagnosis of breast cancer
XX Claim 1; Page 2476; 3695pp; English.
XX The invention relates to human breast cancer expressed polynucleotides

CC (AA107544-AA126789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.

XX Sequence 720 BP: 198 A; 163 C; 115 G; 244 T; 0 other;

Query Match 10.7%; Score 32; DB 22; Length 720;
Best Local Similarity 53.1%; Pred. No. 0.87;
Matches 68; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

OY 57 AATACATCATCATTTACCTGCGACTGTGGGCGCTATCGCACTTAACGTTTCGTCGCC 116
DB 594 AATCCCATCTCTGAGAAATTTGATTTGGTCTGCTGAGAGAAATTAAGTAATTTGTC 535
OY 117 CCCGACGCTATGCAATAGACCAATTAAGTCAAAAGTCCGCTGATAGGCTGAAA 176
DB 534 TGTGAGTTTAAATGAATTTGAATTTAATGAAGGAACATATGCTTTATTTTGAA 475
OY 177 AGTTCATT 184
DB 474 AGGACATT 467

RESULT 8

AA584698
ID AA584698 standard; cDNA: 765 BP.

XX AA584698;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #20502.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX P-PSDB: ABG20511.

XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity

XX Claim 1; SEQ ID NO 20502; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome

XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags

XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving

(II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations and to produce other types of data and products dependent on DNA and amino acid sequences. MAS64197-MAS94564 represent novel human diagnostic coding sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at http://wipo.int/pub/published_pcl_sequences.

Sequence 765 BP; 250 A; 191 C; 185 G; 139 T; 0 other;

Query Match 10.3%; Score 30.8; DB 23; Length 765;
Best Local Similarity 57.1%; Pred. No. 2.4;
Matches 56; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 121 CAGTCTATGCAATGACCAATTAACGCAAAAAAGTCGCTGATTAAGCCTTGAAGTT 180
DB 422 CAGCTTACCACCAATGAGACGAAACCAAAATTCGGGAATATGATGAACAAGTT 481
QY 181 CATTCCAGACCATTTTACATCTGACCCGATGAGA 218
DB 482 CTTTACCAACCCCAAGATGACACTGCTCATGAGCA 519

RESULT 9
ABLA9512
ID ABLA9512 standard; cDNA; 1252 BP.
AC ABLA9512;

01-JUL-2002 (first entry)

Sequence #114 used to create transgenic plant with dwarf phenotype.

Transgenic plant; dwarf phenotype; ss.

Nicotiana benthamiana.

WO200208411-A2.

31-JAN-2002.

20-JUL-2001; 2001WO-US23315.

20-JUL-2000; 2000US-219943P.

(LARG-) LARGE SCALE BIOLOGY CORP.

Pogue GP, Della-Cioppa GR, Wolfe GM, Zheng W;

WPI; 2002-195875/25.

Creating transgenic plants exhibiting a dwarf phenotype useful in biopharmaceutical manufacturing, comprises expressing in the plant a gene for dwarfism or stunting having a sequence selected from 122 polynucleotide sequences -

Claim 1; Page 174; 178pp; English.

This invention relates to the creation of a transfected or transgenic plant selected from ornamental, horticultural, forestry, medicinal and Nicotiana sp. plants, all of which exhibit a dwarf phenotype by expressing in the plant a DNA comprising a sequence selected from 122 polynucleotide sequences. This sequence represents on of those polynucleotide sequences. The nucleic acid and amino acid sequences are useful for producing transgenic plants exhibiting a dwarf phenotype for use in biopharmaceutical manufacturing. The cDNA sequences may be used

CC in recombinant DNA molecules to direct expression of polypeptides in appropriate host cells.

Sequence 1252 BP; 322 A; 267 C; 316 G; 347 T; 0 other;

Query Match 10.2%; Score 30.6; DB 24; Length 1252;
Best Local Similarity 50.3%; Pred. No. 3.5;
Matches 75; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 29 TCATGAGTTATGACGATCCCTGCAAAATATCATCATCCTCGACCTGCGG 88
DB 1103 TTACGAGTTTTCCTTCTTCATCATCAAAATAGTTTCAAGCTTTCATCTTTGGG 1162
QY 89 CTATGCCACTTATACCTTTGCTGCTCCCTCAGTCTATGCAATGACCAATTAAGTGA 148
DB 1163 TTTCCGATTTTATGTTGTTCTTGTATGATGAAGACTATGATTAAGCAGCTTATTA 1222
QY 149 AAAAAAGTCGCTGATTAAGCCTTGA 177
DB 1223 GTATGCTTCTTCTTAAAAA 1251

RESULT 10
ABK30708
ID ABK30708 standard; cDNA; 1252 BP.
AC ABK30708;

23-APR-2002 (first entry)

Plant dwarfing/stunting related cDNA seq ID 122.

Plant; ss; dwarfism; stunting; EST; expressed sequence tag;

transgenic plant; plant metabolism.

Nicotiana benthamiana.

WO200208410-A2.

31-JAN-2002.

20-JUL-2001; 2001WO-US23120.

20-JUL-2000; 2000US-219809P.

20-JUL-2000; 2000US-219810P.

(DOMC) DOW CHEM CO.

(REDD) REDDY S A.

(LARR) LARRINUA M I.

(RUEG) RUEGGER M.

(WEG) WEGLARZ T.

(BLAK) BLAKESLEE B.

(ORIE) ORIEO V B J.

(SAVI) SAVICKAS J P.

(MCCR) MCCREARY A D.

(MILL) MILLER A B.

(GACH) GACHOTTE D.

(GROS) GROSLEY R.

(PELL) PELL R.

Reddy SA, Larinna MI, Ruegger M, Weglarz T, Blakeslee B;

Orieo VJ, Savickas JP, McCreary AD, Miller AB, Pogue PG;

Della-Cioppa GR, Wolfe GM, Zheng W, Gachotte D, Grosley R, Pell R;

WPI; 2002-164823/21.

Polynucleotide and amino acid sequences identified in one or more metabolic pathways that lead to dwarfism and stunting in plants, useful in agriculture to create dwarf varieties of any plant species -

Example 10; Fig 1; 717pp; English.

The invention relates to polynucleotide and amino acid sequences

identified in one or more metabolic pathways that lead to dwarfism and stunting in plants. Also included are vectors comprising the polynucleotides, transgenic plants (including the seed and leaf) transfected with the polynucleotides or vectors, a process for altering the metabolism of a plant comprising providing the above vector and a plant, and transfecting the plant with the vector under conditions such that the metabolism of the plant is altered by expression of the isolated nucleic acid from the vector, e.g. such that a stunting phenotype in an industrial plant is produced and a process for the characterization of fractionated biological samples, comprising (a) providing one or more apparatus, a mass spectroscopy apparatus or data analysis software and (b) treating the fractionated biological samples and the reference samples with the gas chromatography apparatus to generate chromatographic data corresponding to the fractionated biological samples and the reference samples, (c) treating the mass spectroscopy apparatus to generate spectroscopic data corresponding to the fractionated biological samples and the reference samples and (d) processing the chromatographic and the spectroscopic data with the data analysis software. The nucleic acid and the vector are useful for altering the metabolism of a plant and for stunting a plant. The nucleic acids are useful in agriculture to create dwarf varieties of any plant species. The present sequence is a plant EST, expressed sequence tag, isolated from a cDNA library and related to stunting/dwarfism.

Sequence 1252 BP: 322 A; 267 C; 316 G; 347 T; 0 other;

Query Match 10.2%; Score 30.6; DB 24; Length 1252;

Best Local Similarity 50.3%; Pred. No. 3.5; Mismatches 75; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

29 TGATGAGTTATGACGATCCCTGCTGAATAATACATCATTCATTCGACGCTGCGCG 88
 1103 TTACGAGTTTCTCTTCAATCATCAATAATAGTTTCAGCTTCACTTCTTGGGG 1162
 89 CTATCGACTTAACTGATGCTGCTGCGCCCTGAGTCTATGACATAGACCTAACTGCA 148
 1163 TTTCGAGTTTAAATGAGTTTCTTGTATGATGAAAGACTATGACATTAAGCACTTAATTA 1222
 149 AAAAAAGTCGCTGATAGCGCTGAAAA 177
 1223 GTAGCTTCTCTCTTAATAAAAAA 1251

RESULT 11
 AAD00097 standard; cDNA; 1584 BP.

AC AAD00097;
 DT 31-JUL-2000 (first entry)
 XX Murine zsig49 cDNA.
 XX Murine zsig49 cDNA; metabolic disease; diabetes; pancreatic disease;
 KW reproductive development; testicular function; wound healing;
 KW anti-microbial; antidiabetic; gene therapy; ss.
 XX Mus musculus.

XX Key Location/Qualifiers
 FH 1..1386
 FT CDS
 FT sig_peptide
 FT 1..81
 FT /*tag= a
 FT /*product= "Murine zsig49 protein"
 FT 82..1383
 FT /*tag= b
 FT mat_peptide
 FT /*tag= c
 FT /*product= "Mature murine zsig49 protein"
 XX WO200023591-A2.

PD 27-APR-2000.
 XX 20-OCT-1999; 99MO-US24579.
 PF 21-OCT-1998; 98US-0176545.
 PR (ZYMO) ZYMOGENETICS INC.

PI Sheppard PO, Holly RD, Gao Z, Whitmore TE, Maurer MF;
 PI WPI: 2000-339690/29.
 DR P-PSDB: AAY70815.

PT zsig49 polypeptides and polynucleotides, useful for treatment and
 PT diagnosis of metabolic diseases such as diabetes -
 PS Claim 19; Page 126-128; 133pp: English.

The present cDNA sequence encodes murine zsig49 protein. zsig49 genes and proteins are useful for treatment and diagnosis of metabolic diseases such as diabetes, pancreatic conditions. They may exert regulatory effects on male gametes, reproductive development and testicular functions. They are therapeutically useful for aiding digestion, wound healing and anti-microbial applications.

Sequence 1584 BP: 467 A; 338 C; 374 G; 405 T; 0 other;

Query Match 10.2%; Score 30.6; DB 21; Length 1584;

Best Local Similarity 56.4%; Pred. No. 4; Mismatches 57; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

146 GCAAAAAAAGTCGCTGATAGAGCTTGAAGAAGTTCATTCACAGCCATTTTACATCG 205
 422 GCAGAGACCGTTCAATTTGACTGCGGAGATATGATGATTCGAGACACCTTTGTGAAC 481
 206 TACCCGATGAGAGCGCGCTGATGGGTCTTCTGCTACCTG 246
 482 TTGATGATGAGCGGCTCTCTCTGCTGACAGATATTTGACCTG 522

RESULT 12

AAC25839/C
 ID AAC25839 standard; cDNA; 443 BP.

AC AAC25839;
 DT 06-OCT-2000 (first entry)
 XX Human secreted protein 5' EST, SEQ ID NO: 29914.
 DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 XX Homo sapiens.

PN EP1033401-A2.
 XX 06-SEP-2000.
 PD 21-FEB-2000; 2000EP-0200610.
 PF 26-FEB-1999; 99US-0122487.
 XX (GEST) GENSET.
 XX Dumas Milne Edwards J, Duclert A, Giordano J;
 XX WPI: 2000-500381/45.

New nucleic acid that is a 5' expressed sequence tag (5' EST) for obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for diagnostic, forensic, gene therapy and chromosome mapping procedures -

.62T96T0-S00000Z,0000,0000,0000

PR	19-MAY-2000	2000US-020551
PR	07-JUN-2000	2000US-020946.57
PR	28-JUN-2000	2000US-021486
PR	30-JUN-2000	2000US-021547
PR	07-JUL-2000	2000US-021635
PR	11-JUL-2000	2000US-021680
PR	11-JUL-2000	2000US-021787
PR	14-JUL-2000	2000US-021890
PR	26-JUL-2000	2000US-0220963
PR	26-JUL-2000	2000US-0220964
PR	14-AUG-2000	2000US-0224518
PR	14-AUG-2000	2000US-0224519
PR	14-AUG-2000	2000US-022513
PR	14-AUG-2000	2000US-0225214
PR	14-AUG-2000	2000US-0225666
PR	14-AUG-2000	2000US-0225267
PR	14-AUG-2000	2000US-0225268
PR	14-AUG-2000	2000US-0225447
PR	14-AUG-2000	2000US-0225757
PR	14-AUG-2000	2000US-0225758
PR	14-AUG-2000	2000US-0225759
PR	18-AUG-2000	2000US-0226279
PR	22-AUG-2000	2000US-0226681
PR	22-AUG-2000	2000US-0226866
PR	23-AUG-2000	2000US-0227182
PR	30-AUG-2000	2000US-0227902
PR	01-SEP-2000	2000US-0228287
PR	01-SEP-2000	2000US-0228343
PR	01-SEP-2000	2000US-0229344
PR	05-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229509
PR	06-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231242
PR	08-SEP-2000	2000US-0231243
PR	08-SEP-2000	2000US-0231244
PR	08-SEP-2000	2000US-0231413
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0232080
PR	12-SEP-2000	2000US-0232081
PR	14-SEP-2000	2000US-02321968
PR	14-SEP-2000	2000US-0232397
PR	14-SEP-2000	2000US-0232398
PR	14-SEP-2000	2000US-0232399
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0233063
PR	14-SEP-2000	2000US-0233064
PR	21-SEP-2000	2000US-0233423
PR	21-SEP-2000	2000US-0234274
PR	25-SEP-2000	2000US-0234998
PR	25-SEP-2000	2000US-0234998
PR	27-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235534
PR	29-SEP-2000	2000US-0235637
PR	29-SEP-2000	2000US-0236367
PR	29-SEP-2000	2000US-0236368
PR	29-SEP-2000	2000US-0236370
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239335
PR	13-OCT-2000	2000US-0239337
PR	20-OCT-2000	2000US-0240960

	20-OCT-2000;	2000US-0241221.
PR	20-OCT-2000;	2000US-0241785.
PR	20-OCT-2000;	2000US-0241786.
PR	20-OCT-2000;	2000US-0241787.
PR	20-OCT-2000;	2000US-0241788.
PR	20-OCT-2000;	2000US-0241809.
PR	20-OCT-2000;	2000US-0241826.
PR	01-NOV-2000;	2000US-0244647.
PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.
(HUMA-) HUMAN GENOME SCI INC.		
XX	Rosen CA,	Barash SC, Ruben SM;
XX	P-I	
XX	DR	WPI; 2001-457723/49.
DR	P-P	F-SDB; AAU18642.
XX	Isolated polypeptide for treating, preventing and/or prognosing respiratory disorders related to the lung including lung cancers and also for testing and detection e.g. diagnosis -	
PT		
PT		
PS		
CC	Claim 1; SEQ ID No 101; 507pp; English.	
CC	Sequences AAS29839-AAS29930 represent CDNA molecules, which encode the lung antigen polypeptides and the invention. Lung antigen polypeptides and CC their associated polynucleotides are useful in the diagnosis, treatment	
CC		

and prevention of various types of disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A pathological condition can be determined by detecting the presence or absence of a mutation in a lung antigen polynucleotide. The treatable disorders include autoimmune diseases such as rheumatoid arthritis, hyperproliferative disorders such as neoplasms of the breast or liver, cardiovascular disorders such as cardiac arrest, cerebrovascular disorders such as cerebral ischemia, nervous system disorders such as Alzheimer's disease, infections caused by bacteria, viruses and fungi, ocular disorders such as corneal infection, endocrine disorders such as premature labour and infertility, gastrointestinal disorders such as Crohn's disease, renal disorders such as glomerulonephritis and respiratory disorders such as asthma and pleurisy. The polypeptides can also be used to aid wound healing, to prevent skin aging due to sunburn, to maintain organs before transplantation, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO

Query Match Similarity	10.1%	Score 30.4	DB 22	Length 568
Best Local Similarity	54.5%	Pred. No. 2.9		
Matches 61	Conservative	0	Mismatches 51	Indels 0
				Gaps 0
QY	51 GCTGAAATATACATCATCATATGCGTGCACGTGGCGGCGCTATCGCACCTTAACGTTTCGT	110		
DB	255 GTTAAGACTAGACAGCGCTTGATTTGTTCTTCCGCGCTAAGAAACATGTGCTTCCCT	196		
QY	111 GCTGCCCCCTCAGTCTATGCAATAGACCATTAACCTGCAAAAAAATGCGCT	162		
DB	195 CCGTCCCCCAAGTTCTTGAAATGACCTTTTCTTCCACCAAAAAGACAGCT	144		
RESULT 14				
ABA21158				
ID	ABA21158 standard; DNA; 1579 BP.			
XX	ABA21158;			
AC				
XX				
DT	23-JAN-2002 (first entry)			
XX				
DE	Human nervous system related polynucleotide seq ID NO 13489.			
XX				
XX	Human; nootropic; neuroprotective; cyrostatic; dermatological; virucide;			
KW	immunosuppressive; antihlflammatory; anti-HIV; antibacterial; vulnerary;			
KW	antiparasitosis; antislacking; antianaemic; antiarthritic; cancer;			
KW	antihemematic; hepatoprotective; cerebroprotective; antiinflammatory;			
KW	antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;			
KW	antiparasitic; cardiast; immune disorder; cardiovascular disorder;			
XX	neurological disease; Infection; neurotropic; gene therapy; vaccine; ds			
OS	Homo sapiens.			
XX				
PN	WO200159063-A2.			
XX				
PD	16-ANG-2001.			
XX				
PF	17-JAN-2001; 2001WO-US01334.			
XX				
PR	31-JAN-2000; 2000US-0179065.			
PR	04-FEB-2000; 2000US-0180628.			
PR	24-FEB-2000; 2000US-0184664.			
PR	02-MAR-2000; 2000US-0186350.			
PR	16-MAR-2000; 2000US-0189874.			
PR	17-MAR-2000; 2000US-0190076.			
PR	18-APR-2000; 2000US-0198123.			
PR	19-MAY-2000; 2000US-0205515.			
PR	07-JUN-2000; 2000US-0209467.			
PR	28-JUN-2000; 2000US-0214886.			
PR	30-JUN-2000; 2000US-0215135.			
PR	07-JUL-2000; 2000US-0216647.			
PR	07-JUL-2000; 2000US-0216880.			
PR	11-JUL-2000; 2000US-0217487.			

Sequence 1579 BP; 410 A; 346 C; 353 G; 470 T; 0 other;

51 GCTGAAATACATCATTCAGTCGCACTGTGGCGCTATCGCACTTTAACGTTTCGT 110

111 GCTGCCCTCAGTCTATGCAATAGACCATAACTGCAAAAAAAAAAGTCCGCT 162

XX AAS30161:
AC

Human lung antigen genomic DNA #231.

KW anti-rheumatic
KW cerebroprotec

KW	cedriovascular
KW	fungal infect

XX
XX Homo sapiens

AA PD 02-AUG-2001.

PR 31-JAN-2000;
PR 04-FEB-2000;

PR 16-MAR-2000;
PR 17-MAR-2000;

PR 28-JUN-2000;
DD 30-JUN-2000:

PR	11-JUL-2000
PR	11-JUL-2000

PR 14-AUG-2000

PR	14-AUG-2000	2000US-0225213
PR	14-AUG-2000	2000US-0225214
PR	14-AUG-2000	2000US-0225265
PR	14-AUG-2000	2000US-0225267
PR	14-AUG-2000	2000US-0225268
PR	14-AUG-2000	2000US-0225270
PR	14-AUG-2000	2000US-0225447
PR	14-AUG-2000	2000US-0225557
PR	14-AUG-2000	2000US-0225558
PR	14-AUG-2000	2000US-0225579
PR	14-AUG-2000	2000US-0226681
PR	22-AUG-2000	2000US-0226686
PR	22-AUG-2000	2000US-0227182
PR	22-AUG-2000	2000US-0227009
PR	30-AUG-2000	2000US-0228524
PR	01-SEP-2000	2000US-0229287
PR	01-SEP-2000	2000US-0229343
PR	01-SEP-2000	2000US-0229345
PR	01-SEP-2000	2000US-0229509
PR	05-SEP-2000	2000US-0229513
PR	05-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231242
PR	08-SEP-2000	2000US-0231243
PR	08-SEP-2000	2000US-0231244
PR	08-SEP-2000	2000US-0231413
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0232080
PR	08-SEP-2000	2000US-0232081
PR	12-SEP-2000	2000US-0231968
PR	14-SEP-2000	2000US-0232357
PR	14-SEP-2000	2000US-0232358
PR	14-SEP-2000	2000US-0232359
PR	14-SEP-2000	2000US-0232400
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0233063
PR	14-SEP-2000	2000US-0233064
PR	14-SEP-2000	2000US-0233065
PR	21-SEP-2000	2000US-0234223
PR	21-SEP-2000	2000US-0234274
PR	25-SEP-2000	2000US-0234397
PR	25-SEP-2000	2000US-0234398
PR	25-SEP-2000	2000US-0234398
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235835
PR	29-SEP-2000	2000US-0236327
PR	29-SEP-2000	2000US-0236367
PR	29-SEP-2000	2000US-0236368
PR	29-SEP-2000	2000US-0236369
PR	29-SEP-2000	2000US-0236370
PR	02-OCT-2000	2000US-0236802
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239335
PR	13-OCT-2000	2000US-0239337
PR	20-OCT-2000	2000US-0240960
PR	20-OCT-2000	2000US-0241221
PR	20-OCT-2000	2000US-0241785
PR	20-OCT-2000	2000US-0241786
PR	20-OCT-2000	2000US-0241787
PR	20-OCT-2000	2000US-0241808
PR	20-OCT-2000	2000US-0241809
PR	20-OCT-2000	2000US-0241826
PR	01-NOV-2000	2000US-0244617
PR	08-NOV-2000	2000US-0246474
PR	08-NOV-2000	2000US-0246475
PR	08-NOV-2000	2000US-0246476
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246478

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 17:51:17 ; Search time 54.3956 Seconds
(without alignments)
1691.369 Million cell updates/sec

Title: US-09-990-099-21

Perfect score: 300
Sequence: 1 atcatctgtcgtacgtcgtg9.....gattccagagcattcatt 300

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 segs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/prodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/prodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/prodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/prodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/prodata/1/lna/6C.COMB.seq:*
6: /cgn2_6/prodata/1/lna/6D.COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	29.8	9.9	2437 1 US-08-456-647B-3	Sequence 3, Appl 1
2	29.8	9.9	2437 2 US-08-237-401A-3	Sequence 3, Appl 1
3	29.4	9.8	481 1 US-08-480-784-16	Sequence 16, Appl 1
4	29.4	9.8	481 1 US-08-483-553-16	Sequence 16, Appl 1
5	29.4	9.8	481 1 US-08-487-002-16	Sequence 16, Appl 1
6	29.4	9.8	481 1 US-08-483-554B-16	Sequence 16, Appl 1
7	29.4	9.8	481 1 US-08-488-011B-16	Sequence 16, Appl 1
8	29.4	9.8	481 4 US-08-850-727-16	Sequence 16, Appl 1
9	29.4	9.8	481 5 PCT-US95-10202-16	Sequence 16, Appl 1
10	29.4	9.8	481 5 PCT-US95-10203-16	Sequence 16, Appl 1
11	29.4	9.8	481 5 PCT-US95-10220-16	Sequence 16, Appl 1
12	29.4	9.8	1701 4 US-09-023-731-2	Sequence 2, Appl 1
13	28.2	9.7	349 4 US-09-060-756-612	Sequence 612, App 1
14	28.6	9.5	6534 4 US-09-194-613-4	Sequence 4, Appl 1
15	28.6	9.5	43804 4 US-09-177-461-1	Sequence 1, Appl 1
16	28.6	9.3	3063 4 US-08-294-312B-3	Sequence 3, Appl 1
17	28.6	9.3	3063 4 US-08-468-024B-3	Sequence 3, Appl 1
18	27.8	9.3	741 4 US-09-641-638-81	Sequence 81, Appl 1
19	27.4	9.1	1298 3 US-08-948-705-3	Sequence 3, Appl 1
20	27.2	9.1	231 1 US-07-691-079C-8	Sequence 8, Appl 1
21	27.2	9.1	1056 4 US-08-433-037-18	Sequence 18, Appl 1
22	27.2	9.1	1056 4 US-09-364-230-27	Sequence 27, Appl 1
23	26.8	8.9	2520 4 US-08-961-527-14	Sequence 14, Appl 1
24	26.8	8.9	2570 4 US-08-448-603A-31	Sequence 31, Appl 1
25	26.8	8.9	2570 4 US-09-134-075-31	Sequence 31, Appl 1
26	26.8	8.9	2570 4 US-09-492-739-31	Sequence 31, Appl 1
27	26.8	8.9	2573 2 US-08-448-603A-29	Sequence 29, Appl 1

C 28	26.8	8.9	2573 3 US-09-134-075-29	Sequence 29, Appl 1
C 29	26.8	8.9	2573 4 US-09-492-739-29	Sequence 29, Appl 1
C 30	26.8	8.9	3671 4 US-08-932-787B-5	Sequence 5, Appl 1
C 31	26.8	8.9	3671 4 US-08-932-012C-5	Sequence 5, Appl 1
C 32	26.8	8.9	3671 4 US-08-888-818C-5	Sequence 5, Appl 1
C 33	26.8	8.9	3767 4 US-08-928-941D-28	Sequence 28, Appl 1
C 34	26.8	8.9	3767 4 US-08-928-941D-30	Sequence 30, Appl 1
C 35	26.8	8.9	3767 4 US-09-280-590A-28	Sequence 28, Appl 1
C 36	26.8	8.9	3767 4 US-09-280-590A-30	Sequence 30, Appl 1
C 37	26.6	8.9	4602 3 US-08-904-452-3	Sequence 3, Appl 1
C 38	26.6	8.9	4602 4 US-09-517-639-3	Sequence 3, Appl 1
C 39	26.4	8.8	949 4 US-08-221-017B-615	Sequence 615, App 1
C 40	26.4	8.8	2137 3 US-08-821-818-4	Sequence 4, Appl 1
C 41	26.4	8.8	24417 2 US-08-846-762-1	Sequence 1, Appl 1
C 42	26.2	8.7	1668 1 US-08-463-090B-1	Sequence 1, Appl 1
C 43	26.2	8.7	13977 4 US-09-484-970B-60	Sequence 60, Appl 1
C 44	26.2	8.7	4403765 4 US-09-103-840A-2	Sequence 2, Appl 1
C 45	26.2	8.7	4411529 4 US-09-103-840A-1	Sequence 1, Appl 1

ALIGNMENTS

RESULT 1
US-08-456-647B-3
Sequence 3, Application US/08456647B

Patent No. 5811516

GENERAL INFORMATION:

APPLICANT: Lemke Ph.D. et al., Greg E.
TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES
NUMBER OF SEQUENCES: 54

CORRESPONDENCE ADDRESS:
ADDRESS: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400

CITY: La Jolla
STATE: CA

COUNTRY: US

ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/456,647B
FILING DATE: 02-JUN-1995

CLASSIFICATION: 530

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/237,401
FILING DATE: 02-MAY-1994

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/884,486
FILING DATE: 15-MAY-1992

ATTORNEY/AGENT INFORMATION:
NAME: Wetherell Ph.D., John R.

REGISTRATION NUMBER: 31,678
REFERENCE/DOCKET NUMBER: 07251/007002

TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 678-5099
TELEFAX: (619) 678-5070

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 2437 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

IMMEDIATE SOURCE:

CLONE: Tyro-2

FEATURE:

NAME/KEY: CDS
LOCATION: 3..2118

US-08-456-647B-3

[illegible]

RESULT 2
US-08-237-401A-3

```

1 sequence 3; APPLICATION US/08237401A
2 Patent No. 5837448
3
4 GENERAL INFORMATION:
5   APPLICANT: Lemke Ph.D. et al., Greg E.
6   TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES
7   NUMBER OF SEQUENCES: 54
8   CORRESPONDENCE ADDRESS:
9     ADDRESSEE: Fish & Richardson P.C.
10    STREET: 4225 Executive Square, Suite 1400
11    CITY: La Jolla
12    STATE: CA
13
14 COUNTRY: US
15 ZIP: 92037
16
17 COMPUTER READABLE FORM:
18   MEDIUM TYPE: Floppy disk
19   COMPUTER: IBM PC compatible
20   OPERATING SYSTEM: PC-DOS/MS-DOS
21   SOFTWARE: Patentin Release #1.0, Version #1.25
22
23 CURRENT APPLICATION DATA:
24   APPLICATION NUMBER: US/08/237,401A
25   FILING DATE: 02-MAY-1994
26   CLASSIFICATION: 435
27
28 PRIOR APPLICATION DATA:
29   APPLICATION NUMBER: US 07/884,486
30   FILING DATE: 15-MAY-1992
31
32 ATTORNEY/AGENT INFORMATION:
33   NAME: Halle Ph.D., Lisa A.
34   REGISTRATION NUMBER: 38,347
35   REFERENCE/DOCKET NUMBER: 07251/007001
36
37 TELECOMMUNICATION INFORMATION:
38   TELEPHONE: (619) 678-5070
39   TELEFAX: (619) 678-5099
40
41 INFORMATION FOR SEQ ID NO: 3:
42   SEQUENCE CHARACTERISTICS:
43     LENGTH: 2437 base pairs
44     TYPE: nucleic acid
45     STRANDEDNESS: single
46     TOPOLOGY: linear
47     MOLECULE TYPE: DNA
48     IMMEDIATE SOURCE:
49       CLONE: Tyro-2
50       FEATURE:
51
52     NAME/KEY: CDS
53     LOCATION: 3..2118
54
55 US-08-237-401A-3

```

	Query Match	9.9%	Score 29.8;	DB 2;	Length 2437;	
	Best Local Similarity	55.2%;	Pred. No. 1.1;			
	Matches	58;	Conservative	0;	Mismatches	47;
					Indels	0;
					Gaps	0;
QY	87	GGCTATCGCAGCTTTACGCTTTCGTCTGCCCCCTCAGTCTATGCATAATGACCATTAACTG	146			
Dd	1557	GACAGTACTACTCAGAGGTATATGTTCTCATCCACAGTGTTGCCCCCAAGAAGCATCCT	1616			
QY	147	CAAAAAAAGTCCCGCTGATTAAGGCTTGAAGAATTCATTTCAGAC	191			
Dd	1617	CGAGAGAACTGGATGAGGAAGGCTACTGCATCCAAATGCATGAC	1661			

RESULT 3
US-08-480-784-16/c
; Sequence 16, Application US/08480784
; Patent No. 5693473
GENERAL INFORMATION

```

1 GENERAL INFORMATION:
2 APPLICANT: SKOLNICK, Mark H.
3 APPLICANT: Goldgar, David E.
4 APPLICANT: Miki, Yoshio
5 APPLICANT: Swenson, Jeff
6 APPLICANT:
7 APPLICANT: Kamb, Alexander
8 APPLICANT: Hartsman, Keith D.
9 APPLICANT: Shattuck-Eidens, Donna M.
10 APPLICANT: Tavtigian, Sean V.
11 APPLICANT: Wiseman, Roger W.
12 APPLICANT: Futreal, P. Andrew
13 TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
14 TITLE OF INVENTION: Susceptibility Gene
15 NUMBER OF SEQUENCES: 85
16 CORRESPONDENCE ADDRESS:
17 ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
18 STREET: 1201 New York Avenue, N.W., Suite 1000
19 CITY: Washington
20 STATE: DC
21 COUNTRY: USA
22 ZIP: 20005
23 COMPUTER READABLE FORM:
24 MEDIUM TYPE: Floppy disk
25 COMPUTER: IBM PC compatible
26 OPERATING SYSTEM: PC-DOS/MS-DOS
27 SOFTWARE: PatentIn Release #1.0, Version #1.30
28 CURRENT APPLICATION DATA:
29 APPLICATION NUMBER: US/08/480,784
30 FILING DATE:
31 CLASSIFICATION: 435
32 PRIOR APPLICATION DATA:
33 APPLICATION NUMBER: US 08/409,305
34 FILING DATE: 24-MAR-1995
35 PRIOR APPLICATION DATA:
36 APPLICATION NUMBER: US 08/348,824
37 FILING DATE: 29-NOV-1994
38 PRIOR APPLICATION DATA:
39 APPLICATION NUMBER: US 08/308,104
40 FILING DATE: 16-SEP-1994
41 PRIOR APPLICATION DATA:
42 APPLICATION NUMBER: US 08/300,266
43 FILING DATE: 02-SEP-1994
44 PRIOR APPLICATION DATA:
45 APPLICATION NUMBER: US 08/289,221
46 FILING DATE: 12-AUG-1994
47 ATTORNEY/AGENT INFORMATION:
48 NAME: Innen, Jeffrey L.
49 REGISTRATION NUMBER: 28, 957
50 REFERENCE/DOCKET NUMBER: 24884-109347
51 TELECOMMUNICATION INFORMATION:
52 TELEPHONE: 202-962-4810
53 TELEFAX: 202-962-8300
54 INFORMATION FOR SEQ ID NO: 16:
55 SEQUENCE CHARACTERISTICS:
56 LENGTH: 481 base pairs
57 TYPE: nucleic acid
58 STRANDEDNESS: double
59 TOPOLOGY: linear
60 MOLECULE TYPE: DNA (genomic)
61 HYPOTHEetical: NO
62 ANTI-SENSE: NO
63 ORIGINAL SOURCE:
64 ORGANISM: Homo sapiens
65 US-08-480-784-16

```

Query Match	9.8%;	Score 29.4;	DB 1;	Length 481;
Best Local Similarity	53.1%;	Pred. No. 0.65;		

Matches 60: Conservative 1; Mismatches 52; Indels 0; Gaps 0;

QY 107 TCGTGTGCCCCCTCAGTCTATGCAATGACCAATGCAAAAAAGTCCGCTGATA 166

DB 122 TCCAGCTGGGACAAACAGAGTAAAGCCCTGTCACAAAAAAGAGTACGTGTG 63

QY 167 AGGCTTGAAGATTCATTTCCAGACCCATTTTACATCGTACCCGATGAGAC 219

DB 62 AAAAATTAAAGTTCATTTATTTGCTCACAATCTCTTTGACCAATGAGTGC 10

RESULT 4

US-08-483-553-16/C

Sequence 16, Application US/08483553

Patent No. 5709999

GENERAL INFORMATION:

APPLICANT: Skolnick, Mark H.

APPLICANT: Goldgar, David E.

APPLICANT: Miki, Yoshio

APPLICANT: Swenson, Jeff

APPLICANT: Kamb, Alexander

APPLICANT: Harshtman, Keith D.

APPLICANT: Shattuck-Eidens, Donna M.

APPLICANT: Tavtigian, Sean V.

APPLICANT: Wiseman, Roger W.

APPLICANT: Futreal, P. Andrew

APPLICANT: Baelejer, Howard & Civiletti, LLP

ADDRESS: 1201 New York Avenue, N.W., Suite 1000

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/483,553

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/409,305

FILING DATE: 24-MAR-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/348,824

FILING DATE: 29-NOV-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/308,104

FILING DATE: 16-SEP-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/300,266

FILING DATE: 02-SEP-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/289,221

FILING DATE: 12-AUG-1994

ATTORNEY/AGENT INFORMATION:

NAME: Ihnen, Jeffrey L.

REGISTRATION NUMBER: 28,957

REFERENCE/DOCKET NUMBER: 24884-109347

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-962-4810

TELEFAX: 202-962-8300

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:

LENGTH: 481 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

US-08-483-553-16

Query Match

Best Local Similarity

Matches 60: Conservative 1; Mismatches 52; Indels 0; Gaps 0;

QY 107 TCGTGTGCCCCCTCAGTCTATGCAATGACCAATGCAAAAAAGTCCGCTGATA 166

DB 122 TCCAGCTGGGACAAACAGAGTAAAGCCCTGTCACAAAAAAGAGTACGTGTG 63

QY 167 AGGCTTGAAGATTCATTTCCAGACCCATTTTACATCGTACCCGATGAGAC 219

DB 62 AAAAATTAAAGTTCATTTATTTGCTCACAATCTCTTTGACCAATGAGTGC 10

RESULT 5

US-08-487-002-16/C

Sequence 16, Application US/08487002

Patent No. 5710001

GENERAL INFORMATION:

APPLICANT: Shattuck-Eidens, Donna M.

APPLICANT: Simard, Jacques

APPLICANT: Eml, Mitsuru

APPLICANT: Nakamura, Yusuke

APPLICANT: Durocher, Francine

APPLICANT: Baelejer, Howard & Civiletti, LLP

ADDRESS: 1201 New York Avenue, N.W., Suite 1000

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/487,002

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/409,305

FILING DATE: 24-MAR-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/348,824

FILING DATE: 29-NOV-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/308,104

FILING DATE: 16-SEP-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/300,266

FILING DATE: 02-SEP-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/289,221

FILING DATE: 12-AUG-1994

ATTORNEY/AGENT INFORMATION:

NAME: Ihnen, Jeffrey L.

REGISTRATION NUMBER: 28,957

REFERENCE/DOCKET NUMBER: 24884-109347

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-962-4810

TELEFAX: 202-962-8300

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:

LENGTH: 481 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

LENGTH: 481 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-487-002-16

Query Match
Best Local Similarity 53.1%; Score 29.4; DB 1; Length 481;
Matches 60; Conservative 1; Mismatches 52; Indels 0; Gaps 0;

QY 107 TCGTGTGCCCCCTCAGTCTATGCATATAGCCATTAACGCAAAAAAGTCGGTGATA 166
DB 122 TCCAGCTGGGACACAGAGTAGMCCCTGTCTCAAAAAAGTAGCTGTG 63
QY 167 AGCCTGAAAGTTCATTTCAGACCCATTTTACATCTGATGCGATGAGAC 219
DB 62 AAAAATTAAAGTTCAATTATTTGCTCACAATCTCTTTGACCATGTGTGTC 10

RESULT 6
US-08-483-554B-16/c
Sequence 16, Application US/08483554B
Patent No. 5747282
GENERAL INFORMATION:

APPLICANT: Skolnick, Mark H.
APPLICANT: Goldgar, David E.
APPLICANT: Miki, Yoshio
APPLICANT: Swenson, Jeff
APPLICANT: Kamb, Alexander
APPLICANT: Harshman, Keith D.
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Tavligian, Sean V.
APPLICANT: Wiseman, Roger W.
APPLICANT: Futreal, P. Andrew
TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,554B
FILING DATE: 07-JUN-1995

CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994

ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 481 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-483-554B-16

Query Match
Best Local Similarity 53.1%; Score 29.4; DB 1; Length 481;
Matches 60; Conservative 1; Mismatches 52; Indels 0; Gaps 0;

QY 107 TCGTGTGCCCCCTCAGTCTATGCATATAGCCATTAACGCAAAAAAGTCGGTGATA 166
DB 122 TCCAGCTGGGACACAGAGTAGMCCCTGTCTCAAAAAAGTAGCTGTG 63
QY 167 AGCCTGAAAGTTCATTTCAGACCCATTTTACATCTGATGCGATGAGAC 219
DB 62 AAAAATTAAAGTTCAATTATTTGCTCACAATCTCTTTGACCATGTGTGTC 10

RESULT 7
US-08-488-011B-16/c
Sequence 16, Application US/08488011B
Patent No. 5753441
GENERAL INFORMATION:

APPLICANT: Skolnick, Mark H.
APPLICANT: Goldgar, David E.
APPLICANT: Miki, Yoshio
APPLICANT: Swenson, Jeff
APPLICANT: Kamb, Alexander
APPLICANT: Harshman, Keith D.
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Tavligian, Sean V.
APPLICANT: Wiseman, Roger W.
APPLICANT: Futreal, P. Andrew
TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,011B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347-09
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-8300
TELEFAX: 202-962-4810
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 481 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-488-0118-16

Query Match 9.8%; Score 29.4; DB 1; Length 481;
Best Local Similarity 53.1%; Pred. No. 0.65;
Matches 60; Conservative 1; Mismatches 52; Indels 0; Gaps 0;
QY 107 TCGTGGTGGCCCTCATCTATGCAATAGACCAATGCAAAAAAGTCCGCTATA 166
DB 122 TCCAGCTGGAGACAGAGTAAGMCCCTGTCTCAAAAAAAGTACTGTTG 63
QY 167 AGGCTGAAAGTTCATTTCCAGACCCATTTTACATCGTAGCCGATGAGAC 219
DB 62 AAAATTAAAGTTCATTTATTTGCTCACAATCTATCTCTTTGACCATGTGTC 10

RESULT 8
US-08-850-727-16/c
Sequence 16, Application US/08850727
GENERAL INFORMATION:
APPLICANT: Skolnick, Mark H.
APPLICANT: Goldgar, David E.
APPLICANT: Miki, Yoshio
APPLICANT: Swenson, Jeff
APPLICANT: Kamb, Alexander
APPLICANT: Harshman, Keith D.
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Tavligian, Sean V.
APPLICANT: Wiseman, Roger W.
APPLICANT: Futreal, P. Andrew
TITLE OF INVENTION: 17q-linked Breast and Ovarian Cancer
TITLE OF INVENTION: Susceptibility Gene
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/850,727
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/483,554
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-8300
TELEFAX: 202-962-4810
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 481 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-850-727-16

Query Match 9.8%; Score 29.4; DB 4; Length 481;
Best Local Similarity 53.1%; Pred. No. 0.65;
Matches 60; Conservative 1; Mismatches 52; Indels 0; Gaps 0;
QY 107 TCGTGGTGGCCCTCATCTATGCAATAGACCAATGCAAAAAAGTCCGCTATA 166
DB 122 TCCAGCTGGAGACAGAGTAAGMCCCTGTCTCAAAAAAAGTACTGTTG 63
QY 167 AGGCTGAAAGTTCATTTCCAGACCCATTTTACATCGTAGCCGATGAGAC 219
DB 62 AAAATTAAAGTTCATTTATTTGCTCACAATCTATCTCTTTGACCATGTGTC 10

RESULT 9
PCT-US95-10202-16/c
Sequence 16, Application PC/TUS9510202
GENERAL INFORMATION:
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Simard, Jacques
APPLICANT: Eml, Mitsuru
APPLICANT: Nakamura, Yusuke
APPLICANT: Durocher, Francine
TITLE OF INVENTION: In Vivo Mutations and Polymorphisms
TITLE OF INVENTION: Susceptibility Gene
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10202
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08-308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 481 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
PCT-US95-10202-16

Query Match
Best Local Similarity 9.8%; Score 29.4; DB 5; Length 481;
Matches 60; Conservative 1; Mismatches 52; Indels 0; Gaps 0;

QY 107 TCGTCTGCCCCCTAGTCTATGACATTAAGCAATAAAGTCCGCTGATA 166
      |||||  |||  |||  |||  |||  |||||  |||||  |||||
DB 122 TCCAGCTGGGACACAGAGTAAGCCCTGTCTCAAAAAAAAAAAAAAGTAGCTGTG 63
      |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 167 AGGCTTGAAAAGTTCATTCCAGACCCATTTTACATGCTGACCGATGAGAC 219
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 62 AAAATTTAAGGTTCATTATTTGCTCACAATTCATCTCTTTGACCATGTGTC 10
      |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 10
PCT-US95-10203-16/c
; Sequence 16, Application PC/TUS9510203
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavligian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
```

```
TITLE OF INVENTION: Susceptibility Gene
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10203
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08-308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 481 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
PCT-US95-10203-16

Query Match
Best Local Similarity 9.8%; Score 29.4; DB 5; Length 481;
Matches 60; Conservative 1; Mismatches 52; Indels 0; Gaps 0;

QY 107 TCGTCTGCCCCCTAGTCTATGACATTAAGCAATAAAGTCCGCTGATA 166
      |||||  |||  |||  |||  |||  |||||  |||||  |||||
DB 122 TCCAGCTGGGACACAGAGTAAGCCCTGTCTCAAAAAAAAAAAAAAGTAGCTGTG 63
      |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 167 AGGCTTGAAAAGTTCATTCCAGACCCATTTTACATGCTGACCGATGAGAC 219
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 62 AAAATTTAAGGTTCATTATTTGCTCACAATTCATCTCTTTGACCATGTGTC 10
      |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 11
PCT-US95-10220-16/c
; Sequence 16, Application PC/TUS9510220
; GENERAL INFORMATION:
```


	Query Match	Score 29.4;	DB 5;	Length 481;
	Best Local Similarity	53.13;	Pred. No. 0.65;	
	Matches 60;	Conservative 1;	Mismatches 52;	Indels 0;
			Gaps	
07	TCGCGCTGCCCTCAGTCTGTGCAATGACCATTAACCTGCAAAAAAGGCTCGTATA	166		


```

NAME/KEY: polyA_site
LOCATION: (21836)
FEATURE:
NAME/KEY: polyA_site
LOCATION: (21882)
FEATURE:
NAME/KEY: misc_feature
LOCATION: (23608)
OTHER INFORMATION: /note= 100k splice acceptor site
FEATURE:
NAME/KEY: misc_feature
LOCATION: (23649)
OTHER INFORMATION: /note= 100k splice acceptor site
FEATURE:
NAME/KEY: gene
LOCATION: (23680)..(27886)
OTHER INFORMATION: /gene: L4
FEATURE:
NAME/KEY: polyA_site
LOCATION: (27920)
FEATURE:
NAME/KEY: misc_feature
LOCATION: (28315)
OTHER INFORMATION: /note= fibre splice acceptor site
FEATURE:
NAME/KEY: misc_feature
LOCATION: (28341)
OTHER INFORMATION: /note= fibre splice acceptor site
FEATURE:
NAME/KEY: gene
LOCATION: (28363)..(31768)
OTHER INFORMATION: /gene: L5
FEATURE:
NAME/KEY: misc_feature
LOCATION: (30511)
OTHER INFORMATION: /gene: L5 /note= fibre splice acceptor site
FEATURE:
NAME/KEY: polyA_site
LOCATION: (31770)
US-09-171-461-1

```

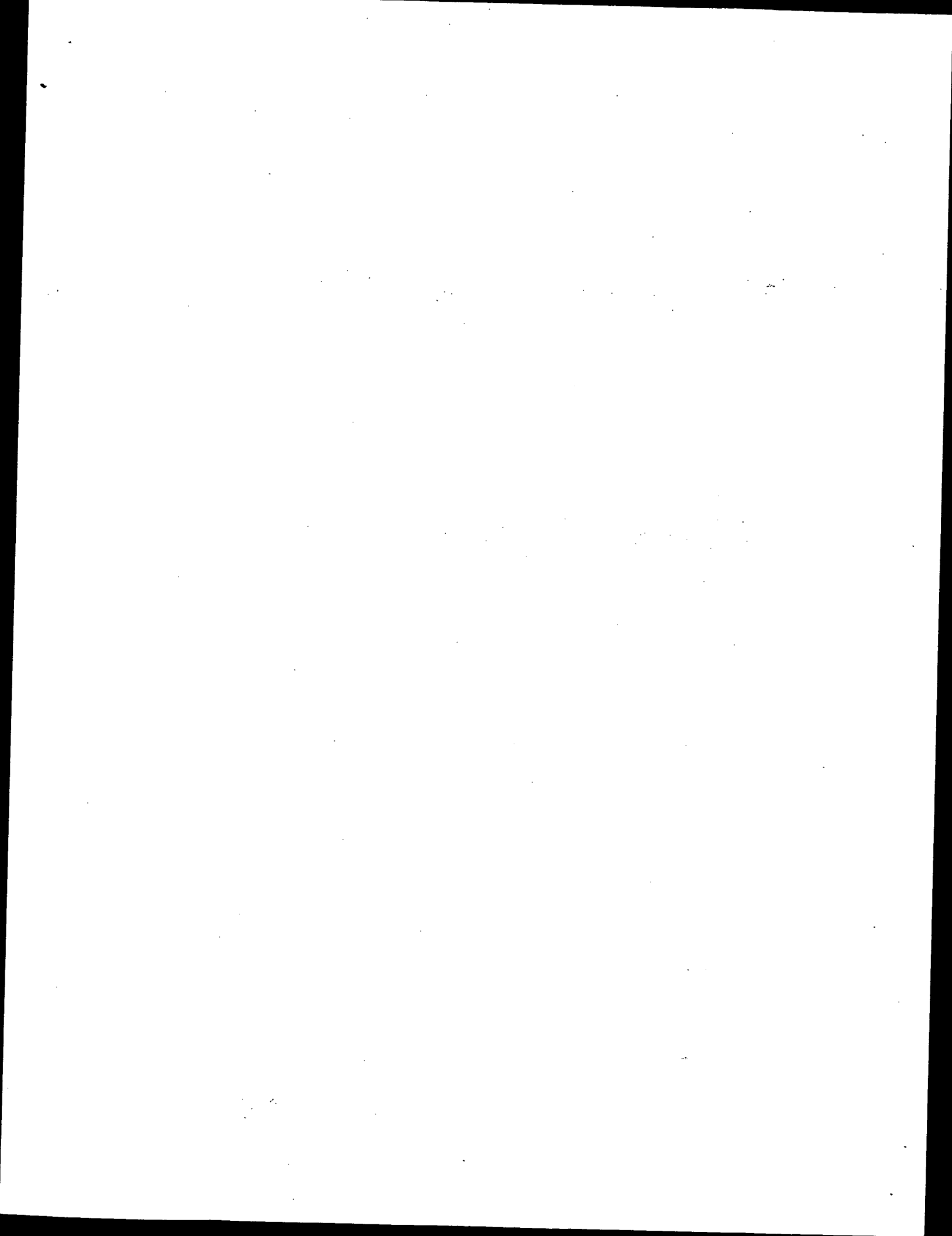
```

Query Match          9.5%; Score 28.4; DB 4; Length 43804;
Best Local Similarity 51.6%; Pred. No. 15;
Matches 65; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

OY 137 CCATTAAGTCAAAAAAGTCCCTGATAGGCTTGAAAGTTCATTCCAGACCCATT 196
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 41222 CCATTAATGCCCATAGGGAATGATGTTCCCTTCAAGTACATTGCCGAATGTA 41163
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 197 TTACATCGTAGCGAGGAGCGGCGCTGATGGGTGTTCTGCTACCTGACCTGTCAT 256
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 41162 CTTAATATGGAGCTAGAGAGAGAGACAGAACAAAGTGTCTGGGGTCCATCGCGCTCGC 41103
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 257 TGTGGA 262
    ||||| |
DB 41102 TGTGTA 41097

```

Search completed: May 5, 2003, 18:00:48
 Job time : 63.3956 secs



RESULT 2
US-09-925-297-225
; Sequence 225, Application US/09925297
; Patent No. US20020081659A1

GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA105
CURRENT APPLICATION NUMBER: US/09/925,297
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05989
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 928
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 225
LENGTH: 1791
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-297-225

Query Match 11.5%; Score 34.4; DB 10; Length 1791;
Best Local Similarity 53.8%; Pred. No. 0.14;
Matches 71; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 153 AAGTCGGCTGATGAGCTTGAAGAAAGTTTCATTTCCAGACCCATTATTACATGTCAGCCCA 212
DB 276 AAGGCTGCTGTTGGCTAGAGAGCCATTATATTCACACATTAAAGACATGAAGC 335
QY 213 TGAGGAGCGCGCTGATGGGTCTGCTGCTACCTGACCTGCTTCATTTGGAAGCTTTACA 272
DB 336 TGTGGAAGACCCCTCGATATTCCTGCAACCCAGAGGTATGTGCTCTCTATCA 395
QY 273 TTTCTGCTGATT 284
DB 396 ACCATTCCAATT 407

RESULT 3
US-10-041-856-1/C
Sequence 1, Application US/10041856
Patent No. US2002016929A1
GENERAL INFORMATION:
APPLICANT: SLAUGENHAUPT, SUSAN
APPLICANT: GUSELLA, JAMES F.
TITLE OF INVENTION: GENE FOR IDENTIFYING INDIVIDUALS WITH FAMILIAL
TITLE OF INVENTION: DYSAUTONOMIA
FILE REFERENCE: 1829-4004US1
CURRENT APPLICATION NUMBER: US/10/041,856
CURRENT FILING DATE: 2002-07-08
PRIOR APPLICATION NUMBER: 60/260,080
PRIOR FILING DATE: 2001-01-06
NUMBER OF SEQ ID NOS: 88
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 66479
TYPE: DNA
ORGANISM: Homo sapiens
US-10-041-856-1

Query Match 11.3%; Score 34; DB 9; Length 66479;
Best Local Similarity 56.1%; Pred. No. 1.2;
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1 ATTCACTGTTGATCGGTGTTGGCTGATGATTATAGCGATCCCTGCTGAATAA 60
DB 50954 ATTATCAATTTATTTATGTTGTTGATTAATTAATTAACCATTCATTCGCAATCA 50895
QY 61 ACATTCATTCATTCGTCGATGCGGCTATGCGCATTTAAGCTTTCGCTG 114
DB 50894 AAGTCATGAAGATTAGCCCTGTTTCTCTAAGAGCTTTATAGTTGCTG 50841

RESULT 4
US-10-003-806-10/C
Sequence 10, Application US/10003806

Patent No. US20020119929A1
GENERAL INFORMATION:
APPLICANT: Bishop, Colin E.
APPLICANT: Agoulrik, Alexander I.
APPLICANT: Zhu, Qichao
TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY
FILE REFERENCE: P02065051/10024824
CURRENT APPLICATION NUMBER: US/10/003,806
CURRENT FILING DATE: 2001-11-02
PRIOR APPLICATION NUMBER: 60/245,872
PRIOR FILING DATE: 2000-11-03
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn version 3.1
SEQ ID NO 10
LENGTH: 173808
TYPE: DNA
ORGANISM: Mouse
US-10-003-806-10

Query Match 10.8%; Score 32.4; DB 12; Length 173808;
Best Local Similarity 62.2%; Pred. No. 7.1;
Matches 51; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 127 ATGCATAGCACCATAACTGCAAAAAAGTCGCTGATAGGCTTGAAGATTTCATTTTC 186
DB 116803 AGGAAAAAGATTAAATACAAAAACAATCAGCCAAAAAGAGATGCAAAAGATGTTCC 116744
QY 187 CAGACCAATTTTACATGCTAG 208
DB 116743 CAGAACAAATGTTACATCATAG 116722

RESULT 5
US-09-910-664-114
Sequence 114, Application US/09910664
Publication No. US2002019464A1
GENERAL INFORMATION:
APPLICANT: POGUE, Greg P.
APPLICANT: DELLA-CIOPPA, Guy R.
APPLICANT: WOLFE, Gerson M.
APPLICANT: ZHENG, Wenjin
TITLE OF INVENTION: METHODS OF CREATING DMARF PHENOTYPES IN
TITLE OF INVENTION: PLANTS
FILE REFERENCE: 00801018900US01
CURRENT APPLICATION NUMBER: US/09/910,664
CURRENT FILING DATE: 2001-07-20
NUMBER OF SEQ ID NOS: 122
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 114
LENGTH: 1252
TYPE: DNA
ORGANISM: Nicotiana benthamiana
US-09-910-664-114

Query Match 10.2%; Score 30.6; DB 9; Length 1252;
Best Local Similarity 50.3%; Pred. No. 2.5;
Matches 75; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 29 TCATAGATTATAGCATCCCTGCTGAATAATACATCATCATTCAGTCGTCGCGG 88
DB 1103 TTACAGATTTTCCCTTCAATCAATCAAAATAGTTTCAAGCTTTCATTTTCCTTTG 1162
QY 89 CTATGCGCATTTAAGCTTTCGCTGCGCCCTCATGCTATGCAATAGACATAAAGTCA 148
DB 1163 TTTGCGATTTTAATGTTTCTTGTGAATGAAGAAGACTATGCAATTAAGGACATTAATA 1222
QY 149 AAAAAAGTCCGCTGATAGAGCTTGAAAA 177
DB 1223 GTAAAGCTTTCTTCTTAATAAAAAAAAAA 1251

RESULT 6
US-09-923-876-2531

Query Match	10.18;	Score 30.4;	DB 5;	Length 50;
Best Local Similarity	54.58;	Pred. No. 1.9;		
Matches	61;	Mismatches	51;	Gaps 0;
	Conservative			

```

RESULT 9
US-10-079-854-425/c
Sequence 425, Application US/10079854
Publication No. US20030054368A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PA121C1
CURRENT APPLICATION NUMBER: US/10/079,854
CURRENT FILING DATE: 2002-02-22
Prior Application removed - See File Wrapper or Palm
NUMBER OF SPO ID NOS: 428
SOFTWARE: PatentIn Ver. 2.0
SPO ID NO: 425
LENGTH: 1579
TYPE: DNA
ORGANISM: Homo sapiens
US-10-079-854-425

```

Query Match

Best Local Similarity 10.1%; Score 30.4; DB 9; Length 1579;
Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 51 GCTGAATAATACATCATTCATTCATTCGCGACGTGGCGGCTATCGCATTTAACTTTCGT 110
DB 590 GTTAAGACTAGCAGCGTGGATTATTTGCTTCTGCTAAGAAACATTTGCTTCCCT 531
QY 111 GCTGCCCCCTCAGTCTATGCAATAGCACTAACTGCAAAAAAAGTCCGCT 162
DB 530 CCTGCCCCCAAGTTCTTGAATGACCTTTCTTTCACCAAAAAAGACACT 479

RESULT 10

US-09-764-878-425/c
; Sequence 425, Application US/09764878
; Patent No. US20020090615A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; CURRENT APPLICATION NUMBER: US/09/764,878
; PRIOR FILING DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 425
; LENGTH: 1579
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-878-425

Query Match

Best Local Similarity 10.1%; Score 30.4; DB 10; Length 1579;
Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 51 GCTGAATAATACATCATTCATTCATTCGCGACGTGGCGGCTATCGCATTTAACTTTCGT 110
DB 590 GTTAAGACTAGCAGCGTGGATTATTTGCTTCTGCTAAGAAACATTTGCTTCCCT 531
QY 111 GCTGCCCCCTCAGTCTATGCAATAGCACTAACTGCAAAAAAAGTCCGCT 162
DB 530 CCTGCCCCCAAGTTCTTGAATGACCTTTCTTTCACCAAAAAAGACACT 479

RESULT 11

US-10-079-854-422/c
; Sequence 422, Application US/10079854
; Publication No. US20030054368A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; CURRENT APPLICATION NUMBER: US/10/079,854
; PRIOR FILING DATE: 2002-02-22
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 422
; LENGTH: 4607
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-079-854-422

Query Match

Best Local Similarity 10.1%; Score 30.4; DB 9; Length 4607;
Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 51 GCTGAATAATACATCATTCATTCATTCGCGACGTGGCGGCTATCGCATTTAACTTTCGT 110
DB 590 GTTAAGACTAGCAGCGTGGATTATTTGCTTCTGCTAAGAAACATTTGCTTCCCT 531

QY 111 GCTGCCCCCTCAGTCTATGCAATAGCACTAACTGCAAAAAAAGTCCGCT 162
DB 530 CCTGCCCCCAAGTTCTTGAATGACCTTTCTTTCACCAAAAAAGACACT 479

RESULT 12

US-10-079-854-424/c
; Sequence 424, Application US/10079854
; Publication No. US20030054368A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; CURRENT APPLICATION NUMBER: US/10/079,854
; PRIOR FILING DATE: 2002-02-22
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 424
; LENGTH: 4607
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-079-854-424

Query Match

Best Local Similarity 10.1%; Score 30.4; DB 9; Length 4607;
Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 51 GCTGAATAATACATCATTCATTCATTCGCGACGTGGCGGCTATCGCATTTAACTTTCGT 110
DB 590 GTTAAGACTAGCAGCGTGGATTATTTGCTTCTGCTAAGAAACATTTGCTTCCCT 531
QY 111 GCTGCCCCCTCAGTCTATGCAATAGCACTAACTGCAAAAAAAGTCCGCT 162
DB 530 CCTGCCCCCAAGTTCTTGAATGACCTTTCTTTCACCAAAAAAGACACT 479

RESULT 13

US-09-764-878-422/c
; Sequence 422, Application US/09764878
; Patent No. US20020090615A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; CURRENT APPLICATION NUMBER: US/09/764,878
; PRIOR FILING DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 422
; LENGTH: 4607
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-878-422

Query Match

Best Local Similarity 10.1%; Score 30.4; DB 10; Length 4607;
Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 51 GCTGAATAATACATCATTCATTCATTCGCGACGTGGCGGCTATCGCATTTAACTTTCGT 110
DB 590 GTTAAGACTAGCAGCGTGGATTATTTGCTTCTGCTAAGAAACATTTGCTTCCCT 531
QY 111 GCTGCCCCCTCAGTCTATGCAATAGCACTAACTGCAAAAAAAGTCCGCT 162
DB 530 CCTGCCCCCAAGTTCTTGAATGACCTTTCTTTCACCAAAAAAGACACT 479

RESULT 14

US-09-764-878-424/c
; Sequence 424, Application US/09764878
; Patent No. US20020090615A1


```
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA121
; CURRENT APPLICATION NUMBER: US/09/764,878
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 424
; LENGTH: 4607
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-764-878-424

Query Match          10.1%; Score 30.4; DB 10; Length 4607;
Best Local Similarity 54.5%; Pred. No. 5.7;
Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

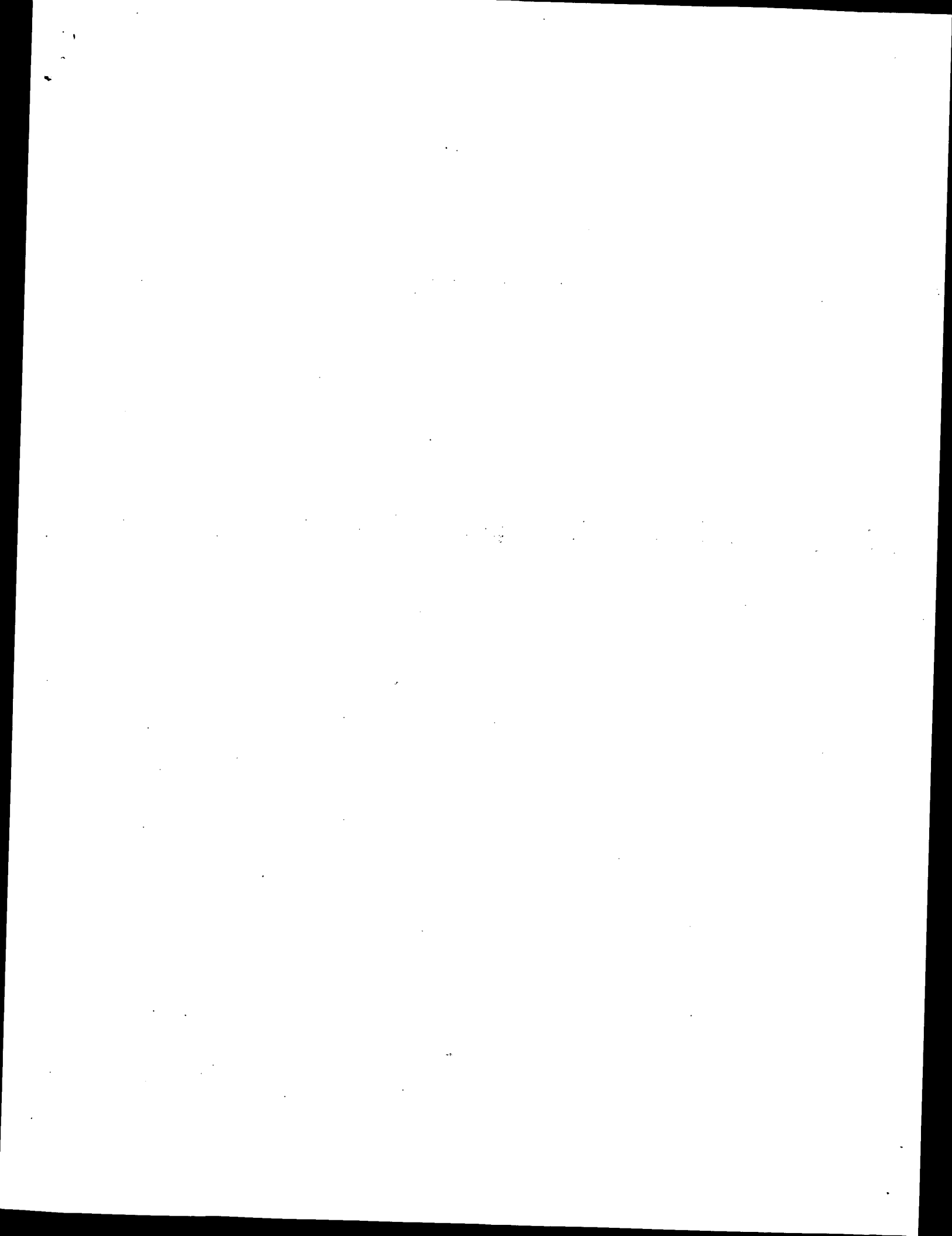
QY 51 GCTGAATAATACATCATCATTAGCTGCCACTGTGGGGCTATGCACTTTAAGTTTGGT 110
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 590 GTTAAGACTAGCAGCGCTTGATTTGTTCTTCTCTGCTAGAAACATGTGCTTCTCT 531
QY 111 GCTGCCCCCTCAGCTATGCAATAGACCATTAACGCAAAAAAAGTCGCT 162
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 530 CTTGCCCCCAAGTTCTTGAATGACACCTTTCTTCACCAAAAAAGACGCT 479

RESULT 15
US-09-901-136-3
; Sequence 3, Application US/09901136
; Publication No. US2003003968A1
; GENERAL INFORMATION:
; APPLICANT: HU, Song et al.
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS, AND
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001273
; CURRENT APPLICATION NUMBER: US/09/901,136
; CURRENT FILING DATE: 2001-07-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 378361
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(378361)
; OTHER INFORMATION: n = A,T,C or G
; US-09-901-136-3

Query Match          10.1%; Score 30.4; DB 9; Length 378361;
Best Local Similarity 57.3%; Pred. No. 53;
Matches 55; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 191 CCCATTTTACATGTCAGCGATGAGACGGCCCTGATGGGTGCTACCTGACCT 250
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 188949 CTCACCTCAGCTGTTACCTTTGTACGTAGATAATCTTTGTTGTGGGCTGTCT 189008
QY 251 GTCCATTGTGGAAGGCTTACATTCGCGTATTC 286
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 189009 GTGCAATTGAGGATGTTAGCAGCATCCCTGATCTC 189044

Search completed: May 5, 2003, 18:05:29
Job time : 245.637 secs
```



GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 17:53:33 ; Search time 1830.49 Seconds
(without alignments)
2654.281 Million cell updates/sec

Title: US-09-990-099-21

Perfect score: 300
Sequence: 1 attcatctgttgcctgctgg.....gattcagagcctatgatt 300

Scoring table: IDENTITY_NUC
Gapop 10.0 ; Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database:

EST:
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estlu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	131.4	43.8	965	17	AF035988
C 2	92.6	30.9	863	17	AF094932
C 3	36.4	12.1	313	10	BB791480
C 4	36	12.0	690	10	BE298245
C 5	35.6	11.9	968	17	CNS03VR3
C 6	35.2	11.7	829	13	BI763074

7	34.4	11.5	373	9	AA336439
8	34.4	11.5	521	12	BF727196
9	34.4	11.5	557	10	BE386483
10	34.4	11.5	623	13	BI869063
11	34.4	11.5	728	9	AL529169
12	34.4	11.5	769	12	BG680800
13	34.4	11.5	779	13	BI764142
14	34.4	11.5	827	13	BI819186
15	34.4	11.5	866	9	AU117405
16	34.4	11.5	872	9	AL530997
17	34.4	11.5	900	14	BQ232193
18	34.4	11.5	952	13	BI517912
19	34.4	11.5	983	14	BQ868982
20	34.4	11.5	1038	9	AL545606
21	34.4	11.5	1063	14	BM905328
22	34.4	11.5	1105	14	BM928240
23	34.4	11.5	1117	14	BM924731
24	34.4	11.5	1133	14	BM924731
25	34.2	11.4	605	13	BM492118
26	34	11.3	666	17	AZ834122
27	33.8	11.3	351	14	BQ283149
28	33.8	11.3	440	9	AI285962
29	33.8	11.3	453	10	BE363625
30	33.8	11.3	595	10	BE595980
31	33.6	11.2	1084	17	CNS05909
32	33.4	11.1	630	10	AW955927
33	33	11.0	572	17	AO670967
34	32.8	10.9	381	12	BI416919
35	32.8	10.9	387	12	BF900486
36	32.8	10.9	502	10	AM602876
37	32.8	10.9	515	13	BM212354
38	32.8	10.9	525	10	AV915911
39	32.8	10.9	585	14	BQ291456
40	32.8	10.9	666	10	AV915809
41	32.8	10.9	681	10	BE438175
42	32.8	10.9	1011	17	CNS04CVP
43	32.8	10.9	1050	17	CNS020USI
44	32.6	10.8	511	17	AZ873575
45	32.4	10.8	575	14	BQ634660

ALIGNMENTS

RESULT 1
AF035988/c 965 bp DNA linear GSS 29-AUG-2000
LOCUS AF035988 Salmonella typhimurium LT2, Lambda DASH II Salmonella
DEFINITION typhimurium genomic clone 7-T7, DNA sequence.

ACCESSION AF035988
VERSION AF035988.1 GI:2689059
KEYWORDS GSS.
SOURCE Salmonella typhimurium.
ORGANISM Salmonella typhimurium
Bacteria; Proteobacteria; gamma subphylum; Enterobacteriaceae;

REFERENCE 1 (bases 1 to 965)
AUTHORS Wong, R.M.-Y., Wong, K.K., Benson, N.R. and McClelland, M.
TITLE Sample sequencing of a Salmonella typhimurium LT2 lambda library:
comparison to the Escherichia coli K12 genome

JOURNAL FEBS Microbiol. Lett. 173 (2), 411-423 (1999)
MEDLINE 99243757
COMMENT Contact: McClelland M
Molecular Biology
Sidney Kimmel Cancer Center
3099 Science Park Road, San Diego, CA 92121, USA
Email: mclelland@fscs.sdsu.edu
Class: Shotgun.

FEATURES

source Location/Qualifiers
1..965
/organism="Salmonella typhimurium"
/strain="LT2"
/db_xref="taxon:602"

(CONTINUATION OF FD-302) (DATE-TIME-CALL 1/31 WEHL 164),

/note- pooled cell lines ; (cell_line=CRL-1751 WEHI 164).

Db 724 GCTTTGATCATTTATCTGCAGAAATTAACATGACTGATTAATTAATTGTTCTGTTCATTTCAG 665

Qy 189 GACCCATTTTACATCGTAGCCGATGAGAGCGCCCTG 226

Db 664 GCTTGTAAATACATCTGATGATGAGAGCGCTGTGTG 627

RESULT 6	BI763074	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
	BI763074	603047876p1	NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5187906 5',	829 bp	linear	EST 25-SEP-2001		
			mRNA sequence.					
	BI763074							
	BI763074.1	GI:15754652						
			EST.					
	human.							
	Homo sapiens							
	Chordata; Craniata; Vertebrata; Euteleostomi;							

```

FEATURES
    source
        Location/Qualifiers
            1..829
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_lib="NIH_MGC_116"
                /shost="DH10B"
                /note="Organ: pooled colon, kidney, stomach; Vector:
                PCMV-SPORT6; Site:1: NotI; site_2: EcoRV (destroyed); RNA
                source anonymous pool of 3 colons, age 26 yo male, 49 yo
                female, 71 yo male colon; 46 yo male kidney, and pool of 2
                stomachs, 62 yo male and 70 yo female. Library is
                oligo-dT primed and directionally cloned (EcoRV site is
                destroyed upon cloning). Average insert size 1.4 kb,
                insert size range 1-3 kb. Library is normalized and
                enriched for full-length clones and was constructed by C.
                Gruber (Invitrogen). Research Genetics tracking code
                023. Note: this is a NIH_MGC Library."
            225 a
            225 c
            179 g
            206 t

```

RESULT 7

AA336439	373 bp	mRNA	linear	EST 21-APR-1997
LOCUS				
DEFINITION	AA336439	EST41018	Endometrial tumor Homo sapiens	cdna 5' end, mRNA sequence
ACCESSION	AA336439			
VERSION	AA336439.1	GI:1988677		
KEYWORDS	EST.			
SOURCE	human.			

The Institute for Genome Sciences and Policy
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org
 For clone availability, additional sequence and expression
 information related to this EST, please check the TIGR Human Gene
 Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..373
 /organism="Homo sapiens"
 /db_xref="ATCC (inost):138136"
 /db_xref="taxon:9606"
 /clone_lib="Endometrial tumor"
 /sex="female"
 /dev_stage="adult"
 /note="Organ: endometrium; Vector: pBluescript SK-;
 Site_1: EcoRI. Site_2: XhoI"
 BASE COUNT 104 a 86 g 90 t 2 others
 ORIGIN
 Query Match 11.5%; Score 34.4; DB 9; Length 373;
 Best Local Similarity 56.9%; Pred. No. 3;
 Matches 62; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

RESULT 8				
BF727196				
LOCUS	BF727196	521 bp	mRNA	linear
				EST 05-JAN-2001

Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LLM12033 row: h column: 18
 High quality sequence stop: 622.
 Location/Qualifiers

FEATURES
 source

1..623
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5405249"
 /clone_lib="NIH_MGC_90"
 /tissue_type="adenocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: liver; Vector: PCMV-SPORE6; Site: 1: NotI;
 Site: 2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 1.7 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

BASE COUNT 178 a 161 c 138 g 146 t

ORIGIN

Query Match 11.5%; Score 34.4; DB 13; Length 623;
 Best Local Similarity 53.8%; Pred. No. 3.5;
 Matches 71; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

OY 153 AAGTCCGCTGATAGGCTGAAAGTTCATTCAGACCCATTTTACATCGTAGCCGA 212
 |||||
 DB 108 AAGGCTGCTGTTGCTTGAAGAGCTTATATATACACATTAAGACATGAGC 167
 |||||
 OY 213 TGAGACGCGCCGATGATGCTGCTGCTACCTGACCTGCTGAGAGCTTACA 272
 |||||
 DB 168 TGTGAAGACCCCTGATGATTCCTGCAACACACAGGCTATGCTCTCTATCA 227
 |||||
 OY 273 TTCTCGCTGATT 284
 |||||
 DB 228 ACCATTCCAAAT 239

RESULT 11
 AL529169 728 bp mRNA linear EST 13-FEB-2001

LOCUS AL529169 LTL_NFL001_NBC4 Homo sapiens CDNA clone CSDD002YE09 5

DEFINITION Prime, mRNA sequence.

ACCESSION AL529169

VERSION AL529169.1 GI:12792662

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 728)

AUTHORS Li, W. B., Gruber, C., Jesse, J. and Polayes, D.

TITLE Full-length CDNA libraries and normalization

JOURNAL Unpublished (2001)

CONTACT Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr.

Location/Qualifiers

1..728

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CSDD002YE09"

/clone_lib="LTL_NFL001_NBC4"

/sex="male"

/tissue_type="neuroblastoma cells"

/lab_host="DH10B"

/note="Organ: brain; Vector: PCMVSPORT 6; 1st strand cDNA

was primed with a NotI-oligo(dT) primer. Five prime end
 enriched, double-stranded cDNA was digested with Not I and
 cloned into the Not I and Eco RV sites of the PCMVSPORT 6
 vector. Library was normalized. Library was constructed
 by Life Technologies. Contact: Feng Liang Life
 Technologies, a division of Invitrogen 9800 Medical Center
 Drive Rockville, Maryland 20850, USA Fax: (1) 301 610
 8337 Email: fliang@lifestech.com URL:
<http://fulllength.invitrogen.com>

BASE COUNT 175 a 201 c 189 g 163 t

ORIGIN

Query Match 11.5%; Score 34.4; DB 9; Length 728;
 Best Local Similarity 53.8%; Pred. No. 3.7;
 Matches 71; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

OY 153 AAGTCCGCTGATAGGCTGAAAGTTCATTCAGACCCATTTTACATCGTAGCCGA 212
 |||||
 DB 398 AAGGCTGCTGTTGCTTGAAGAGCTTATATATACACATTAAGACATGAGC 457
 |||||
 OY 213 TGAGACGCGCCGATGATGCTGCTGCTACCTGACCTGCTGAGAGCTTACA 272
 |||||
 DB 458 TGTGAAGACCCCTGATGATTCCTGCAACACACAGGCTATGCTCTCTATCA 517
 |||||
 OY 273 TTCTCGCTGATT 284
 |||||
 DB 518 ACCATTCCAAAT 529

RESULT 12

LOCUS BG680800 769 bp mRNA linear EST 01-MAY-2001

DEFINITION 602628456F1 NCI_CGAP_Skn4 Homo sapiens CDNA clone IMAGE:4753415 5'

mRNA sequence.

ACCESSION BG680800

VERSION BG680800.1 GI:13912197

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 769)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

CONTACT: Robert Strausberg, Ph.D.

Email: cgabs-remail.nih.gov

Tissue Procurement: James Cleaver, M.D.

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

<http://image.llnl.gov>

Plate: LLM10612 row: P column: 24

High quality sequence stop: 766.

Location/Qualifiers

1..769

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4753415"

/clone_lib="NCI_CGAP_Skn4"

/tissue_type="squamous cell carcinoma"

/lab_host="DH10B (T1 phage-resistant)"

/note="Organ: skin; Vector: PCMV-SPORE6; Site: 1: NotI;
 Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.5kb. Library constructed by Life
 Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 181 a 209 c 203 g 176 t

ORIGIN

Query Match

Best Local Similarity

11.5%; Score 34.4; DB 12; Length 769;

53.8%; Pred. No. 3.8;

Matches	71	Conservative	0	Mismatches	61	Indels	0	Gaps	0
OY	153	AAAGTCCCGCTGAATAAGGCTTGA	AAAGTTCATTCCAGACCATTTT	TACATCGTACGCCA	212				
Db	405	AAAGGCTCTGGTTTGCTCAGAA	GAGTCATTTATATTTACAA	CAATTAAGA	CATGAC	464			
OY	213	TGAGAGCGCGGCTGATGGGTGTC	TGCGCTACCTGACCTGTCCAT	TGTGGAAGGCTTTACA	272				
Db	465	TGTTGAAACACCTCTCGATATTC	TCTGC	AAACCAACGAGTCAAGTGCTCTCT	CAATCA	524			
OY	273	TTCTCGCTGATT	284						
Db	525	ACCATTC	CAATT	536					

[illegible]

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 779)		NIH-MGC http://mgc.nci.nih.gov/ .		
		National Institutes of Health, Mammalian Gene Collection (MGC)		
		Unpublished (1999)		
		Contact: Robert Strausberg, Ph.D.		

Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL),
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can
 be found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LLM11458 row: d column: 20
 High quality sequence stop: 767.
 Location/Qualifiers

FEATURES
 Location/Qualifiers
 1. 779
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5183587"
 /clone_11b="NIH_MGC_116"
 /lab_host="DH10B"
 note="Organ: pooled colon, kidney, stomach; Vector:
 pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
 source anonymous pool of 3 colons, age 26 yo male, 49 yo
 female, 71 yo male colon; 46 yo male kidney, and pool of 2
 stomachs, 62 yo male and 70 yo female. Library is
 oligo-dT primed and directionally cloned (EcoRV site is
 destroyed upon cloning). Average insert size 1.4 kb,
 insert size range 1-3 kb. Library is normalized and
 enriched for full-length clones and was constructed by C.
 Gruber (Invitrogen). Research Genetics tracking code
 023. Note: this is a NIH_MGC Library."

Query Match	Similarity	Score	DB	Length
Best Local	53.88	Pred. No. 3.8		779
Matches	71	Conservative	0	Mismatches 61; Indels 0; Gaps 0
153	AAAGTCGCCTGATAGGCGCTTGA	AAAAGTTTCATTTCCACACCATT	TTTACATCGAGCGGA	212
424	AAAGCGCTCGTTGGCTCCTAGAG	AGTCATTATTTACACACATTA	AAAGACATGAGC	483
213	TGAGACCGCGCTGATGGGTCTT	CGCTACCTGTCACATTGCGAG	CTTTACA	272

Db	484	TGTTTAAACACCCCTCCGTGATATTCCGTGCACCAACGATCATGTGCTCTCTATCA	543
Oy	273	TTCTGCGTGATT	284
		— — —	
Db	544	ACCATTCACATT	555

RESULT	14
B1819186	
LOCUS	827 bp mRNA
DEFINITION	60036992F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:517975 5' , linear EST 04-OCT-2001
ACCESSION	B1819186 mRNA sequence.
VERSION	B1819186
KEYWORDS	B1819186.1 GI:15930736
SOURCE	EST.
ORGANISM	human.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

kuşar, yavaş, melazoza; choirata; cranata; vertebrata; euteleostomi;
mammalia; eutheria; primates; catarrhini; hominidae; homo.
1 (bases 1 to 827)
NIH-MGC <http://imgc.nci.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
unpublished (1999)
Contact: Robert Strausberg, Ph.D.

tissue procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 plate: LLAM11443 row: j column: 24
 high quality sequence set: 824.
 Location/Qualifiers
 1..827
 FEATURES
 source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5177975"
/clone_lib="NIH_MGC_115"
/lab_host="DH10B"
/notes="Organ: pooled brain, lung, testis; Vector:
pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
source anonymous pool of 6 male brains, age range 23-27; 1
male lung, age 27; and 1 male testis, age 69. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.8 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
021. Note: this is a NIH_MGC library."

```

[illegible]

